Articulatory Settings of French and English Monolinguals

Ian WILSON

University of Aizu
E-mail: wilson@u-aizu.ac.jp

1. Introduction

Articulatory setting (henceforth AS) is the underlying articulatory posture of a language. It is a concept that has interested phoneticians for centuries, but was never instrumentally measured until a study by Gick et al. (2004) using old x-ray movie films of English and French speech. These authors studied AS by looking at inter-speech posture (henceforth ISP) – the position of the motionless articulators during inter-utterance pauses. Their results showed that the ISP for Québécois French is significantly different from that for Canadian English. Their study, however, examined only five speakers of each language and its methodology was constrained by the fact that the data they analyzed was based on existing x-ray movie films with limited spatial resolution and clarity, and they had no control over the linguistic stimuli or how they were presented to the subjects. The purpose of the present research is to partially replicate the study of Gick et al. (2004) using a greater number of speakers of French and English, and using an entirely different methodology that has enabled more measurement precision and has allowed for control over the phonetic context of the ISPs that were analyzed. The measurement tools I have used to establish baseline data for English and French will allow new languages to be studied and systematically compared with these data in the future.

Few researchers have quantitatively measured aspects of AS to say conclusively how it is different for two languages. The biggest impediment to measuring AS is ensuring that one is measuring only AS and not being influenced by the articulation of the language's speech segments (Laver 2000).

There have been a number of studies trying to characterize a given language in terms of its overall acoustic properties. If AS underlies speech, surely its effects must be audible in the speech signal. The most common method of measuring the overall acoustic properties of a language is to measure its long term average spectrum (henceforth LTAS), the average of many instantaneous spectra over a reasonably long speech sample. A number of LTAS studies have found a correlation between language spoken and LTAS for individual bilingual speakers, while other studies have failed to find any correlation (see Bruyninckx, Harmegnies, Llisterri, and Poch-Olivé, 1994 for a brief summary). However, it is not necessarily the case that LTAS data should directly correlate with AS. Laver (2000, p. 40) pointed out that "all calculations of a long-term average (whether of articulatory position, auditory impression, or acoustic spectrum) based on all segments [...] will give obvious inaccuracies." The problem is that LTAS is a measure of the sounds of a language - i.e., it is directly affected by the phonetic context of the speech one is examining and there is no way to distinguish which aspects of the speech signal are based on AS, and which are a reflection of the frequency of specific articulations in the language's phonetic inventory. Laver (1978, p. 11) stated, "no articulatory setting normally applies to every single segment a speaker utters", and he called this property of speech segments *segmental susceptibility*. Laver (1980, p. 21) further wrote, "because the successive segments in the stream of continuous speech vary in their susceptibility to the effect of settings, a setting is audible only on an intermittent basis, and even when audible, varies in its prominence, depending on the susceptibility of the segment currently being uttered." Evidence supporting this comes from Harmegnies, Esling, and Delplancq (1989), who found that not all deliberate changes to voice quality have large effects on LTAS. Thus, although LTAS may provide a kind of spectral signature of a language, LTAS probably does not accurately describe that language's underlying AS.

The existence of a language-specific preparatory posture that one assumes just before speaking, i.e., a language-specific ISP, was tested for and confirmed by Gick et al. (2004). They showed that not only was the ISP for Québécois French different from that for Canadian English, but the accuracy of production of the ISP was as high as that for producing the speech sound [i], consistent with the view that the ISP is a speech target posture. Specifically, Gick et al. found the following differences between the Québécois-French ISP and the Canadian-English ISP: The tongue tip (TT), tongue body (TB), and tongue root (TR) were all farther away from the opposing vocal tract surface in the French group compared to the English group. The upper lip was significantly more protruded in English, but the lower lip was significantly more protruded in French. For both the jaw and the velum, there was no difference between the French ISP and the English ISP. Gick et al. did not measure the tongue dorsum.

There are a number of reasons to replicate the Gick et al. study. First, their study examined only 5 speakers of each language and made crosslinguistic generalizations based on only these 10 speakers. Second, the Gick et al. data being analyzed were existing x-ray movie data (Munhall, Vatikiotis-Bateson, and Tohkura, 1994) with limited spatial resolution and clarity. Another methodological issue with the Gick et al. study is that because the data already existed, Gick et al. had no control over the phonetic content of the stimuli being used and how they were presented to the subjects. The stimuli in the original x-ray study were not designed to balance the phonetic context surrounding ISPs. In addition, they were presented to speakers as a list of sentences to be read, thus increasing the chances of anticipatory coarticulation effects on ISP, just as anticipatory coarticulation effects on the ISP of the jaw were found by Hamlet & Stone (1981). Because of this, the language-specific differences that Gick et al. found might have been due to the phonetic context rather than language-specific properties of the ISP.

Another potential factor that may have influenced the results of the Gick et al. study is the method of statistical analysis they employed. As the experimental unit for statistical comparison in their repeated measures study, they used the data obtained from each individual measurement token produced by each individual subject, and then used the *jackknife* procedure (i.e., verification that the means of every subset of N-1 subjects was distributed in a similar way) to justify this choice. Although using data obtained from each individual measurement token as the experimental unit is a common practice among speech researchers, Max & Onghena (1999, pp. 265-266) point out that these types of analyses are at risk of having the assumption of independent error effects violated. Max & Onghena recommend using one value per measurement location per subject (i.e., the mean measurement value across all of a given subject's productions in all trials) as the experimental unit. They state that "despite the agreement on this issue in the contemporary statistical literature, the potential for violations of the assumption continues to occur rather frequently in studies addressing normal or disordered speech-language-hearing processes." (p. 266) If anything, the choice of statistical method in Gick et al. (2004) would have resulted in a greater number of significant differences being reported than should reasonably be expected.

In the present research, it is unlikely that utterance-specific *anticipatory* coarticulation effects were possible because the subjects could not see the next stimulus until they had had a chance to assume an ISP. Although *anticipatory* effects were unlikely, *carry-over* effects were impossible to eliminate while still being sure the subjects remained in speech mode. So, instead of *eliminating* carry-over effects, the phonetic context of the last syllables uttered was tightly controlled across languages. Hamlet & Stone (1981) tried to eliminate any *carry-over* effect by waiting "a few seconds" before manually presenting the next stimulus for the subject to read. Therefore, they assumed that the jaw went to some intermediate position (perhaps absolute rest position, but this is not made explicit) or simply drifted around before assuming the configuration of the next pre-speech posture. Öhman (1967, p. 43) mentioned EMG "evidence" for "basic speech posture" *following* an utterance as if it was common knowledge among EMG speech researchers. Thus, he implied that one does not simply maintain the posture of the last sound of the previous utterance, but that one actively moves the articulators back to the basic speech posture.

2. Method

2.1. Subjects

Initially, 22 paid, naive speakers with normal dentition provided data, although 7 of these were excluded for various reasons. All had had at least some exposure to a second language - they had studied a foreign language in school. However, all of the subjects considered themselves to be monolingual and had not been exposed to an L2 earlier than age 6.

Of the eight monolingual French subjects, none had had formal schooling in English before age 10. All but one (Subject 14) lived in the province of Quebec at the time of the study, unless they had just moved to Vancouver within that week for a short homestay or temporary summer employment. Subject 14 had been living in Vancouver for about one year, but had been using 60% French in her daily life as a nanny for a bilingual family. Before moving to Vancouver, she used 90% French in her daily life. All the monolingual French subjects had monolingual French parents.

Of the seven monolingual English subjects, only two of them had studied French beyond high school, Subjects 2 and 5. After completing all their English trials, these two subjects were asked to read one French trial each. In a native-listener rating task used for separate research, these two subjects received a French rating of 2.6 and 1.9, respectively, where 5 equalled native-like, and thus were classified as monolinguals. All the monolingual English subjects lived in Vancouver at the time of the study and all used nothing but English in their daily lives.

The mean age of all seven English subjects was 27. The mean age of all eight French subjects was 24. Since all subjects were adults and none had reached old age, their L1 was neither developing nor deteriorating, and therefore the difference in the two groups' mean ages was not considered an issue. Of the monolingual English subjects, four were female and three were male. As for the French, six were female and two were male. Since all data was scaled based on an anatomical measurement (see Section 2.3.2), the slight gender mismatch between the two monolingual groups was not considered significant.

2.2. Apparatus

The main pieces of equipment for collecting data were an ultrasound machine for viewing the movements of the tongue in real time, and an Optotrak (Northern Digital Inc.) 3020 optical tracking system for measuring the 3D positions of the lips, jaw, and head relative to the ultrasound probe. The ultrasound monitor used was an Aloka ProSound SSD-5000 with a UST-9118 endo-vaginal 180° electronic 3-9.0 MHz curved array probe. The Optotrak system used consists of a set of three single-axis CCD cameras, with 11-bit hardware resolution, that tracked the movements of 12 infrared-emitting diodes (markers). The Optotrak hardware was controlled using a Northern Digital software program, Collect (version 2.002), running on a PC (Micron Millennia XKU 333).

Subjects were seated in “the experiment chair”, a modified antique ophthalmic examination chair (American Optical Co., model 507-A) with a 2-cup rear headrest adjusted to contact the base of the skull just above the neck, and a forehead stabilizing head restraint (“head stabilizer”) with two rubber pads which were positioned to be lightly touching the subject’s forehead near the hairline.

Ultrasound images were recorded onto digital videotape using a JVC SR-VS20 Mini DV/S-VHS VCR. Simultaneous audio for these ultrasound recordings was recorded using a Sennheiser MKH 416 P48 super-cardioid short shotgun condenser interference tube microphone. The microphone signal was fed into the VCR via a digital mixing console (Yamaha 01V).

Stimuli were displayed to the subjects on an Apple PowerBook G4 17-inch laptop computer at a distance of about 2.5 metres, and at approximately eye level. The stimuli were presented as Microsoft PowerPoint (version 10.1.0) slides.

2.3. Procedure

2.3.1. Data Collection

When a subject arrived for a data collection session, the procedure was as follows. First the subject was shown the equipment to be used, was told the procedure to be followed, and was given the opportunity to ask any questions. Then after signing ethics forms, the subject was seated in the experiment chair and the headrest, the head stabilizer, and the ultrasound probe were adjusted to the proper height. The subject was then moved to a more comfortable chair where the Optotrak markers were attached to his/her lips and jaw.

In this experiment, 12 Optotrak markers were used. Markers 1 through 4 were all permanently attached to a pair of lensless glasses that were worn by each subject and it was assumed that these markers did not move relative to each other. Marker 3 was on the midsagittal plane and markers 2 and 4 were equidistant from it. Marker 3 was slightly higher and more protruded from the subject’s face than markers 2 and 4. Marker 1 was situated on a rigid bamboo skewer that was mounted off the right arm of the glasses. Bamboo was used because it is strong enough to remain rigid but light enough not to put the glasses off balance. For all subjects, marker 1 was located to the right of, posterior to, and superior to the subject’s right ear. Note that during the course of a trial, if it is assumed that the glasses do not move relative to the subject’s head, then markers 1 through 4 defined a rigid body that included the subject’s skull. This was important for being able to track the movement of the subject’s skull (and thus the palate as well) during a trial.

Markers 5 and 6 were attached to the ultrasound probe, 70 mm and 140 mm, respectively, from the tip of the probe. Marker 7 was mounted on a 1 cm cube of open cell foam that was taped under the chin using 3M Micropore surgical tape. Markers 8 and 10 were placed at the right and left corners, respectively, of the subject's mouth, as close as possible to the mouth opening without making it uncomfortable when closing the mouth. Marker 9 was placed as close as possible to the vermilion border of the upper lip on the midsagittal plane. Marker 11 was also placed on the midsagittal plane, but on the lower lip. Depending on how "pouty" the subject's lower lip was, it was sometimes necessary to place Marker 11 above the vermilion border in order for its light to be seen consistently by the Optotrak position sensor. Marker 12 was left in place on a wooden, hinged clapper between experiments. The clapper provided a sound that was used to synchronize the Optotrak data with the ultrasound data.

Throughout all ultrasound data collection, the forehead stabilizer and ultrasound probe were locked into position. Water-soluble ultrasound gel was applied to the head of the ultrasound transducer, which was then placed against the subject's neck in the submental region. The probe was positioned so that a midsagittal image was being displayed with the tongue tip towards the right side of the screen. The probe angle was adjusted so that the image on the ultrasound monitor showed as wide a tongue region as possible, from the shadow of the hyoid bone on the left to the mandible shadow on the right. The exact angle of the probe was different for every subject, dependent on anatomy and posture.

A preliminary trial was done prior to any of the main trials where the subject was asked to read sentences. This was a 40-second "wag" trial, the purpose of which was to set Optotrak baselines for movements of the head, lips, and jaw. The subject was asked to turn his/her head to the extreme right, left, up and down bringing the head back to a centre position and pausing between each direction. The subject cycled through this order twice. After the head-turning task, the subject was asked to spread the corners of the lips as widely as possible, as if saying an exaggerated [i]. This was followed by the subject protruding the lips as far as possible, as if saying an exaggerated [u]. The subject was specifically asked to spread markers 8 and 10 as far to the sides as possible, and to protrude markers 9 and 11 out as far as possible. This was done twice each. This spreading and protruding of the lips enabled a baseline to be set for the extremes of lip movement of each subject. For the final few seconds of the wag trial, the subject was asked to relax, look straight ahead at the computer screen and keep the jaw and lips closed. The jaw here was not in a clenched position, but instead set a baseline for a maximally elevated rest position of the jaw.

The main trials involved the subjects reading a number of sentences aloud. Due to the fact that the English stimuli contained some nonsense or low frequency words for use in a different study (Campbell, 2004), all English subjects were given a 15-sentence practice trial. The practice trial was not deemed necessary for the

French subjects because the French stimuli all contained standard vocabulary familiar to any French speaker. The French stimuli were chosen for a future study that needs many tense-lax minimal pairs in a carrier sentence.

Each subject read six blocks of stimuli. The duration of Optotrak data collection was 67 seconds for the practice trial, and 131 seconds for each real trial. Each of the real trials consisted of 30 sentences that were displayed one at a time to the subject. The PowerPoint "slide transition" for each 30-sentence trial was set so that each sentence slide was displayed for 3 seconds followed by a blank slide for 1 second. The final blank slide after the 30th sentence was accompanied by a distinct sound (a loon call), indicating to all that the trial was complete. As each sentence was displayed, the computer beeped, thus making a record on the ultrasound DV tape of when the subject saw what he or she was supposed to say next. It was assumed that before the beep, any preparatory vocal tract posture (see Schmidt and Lee (1999, pp. 126-127) for a description of "preparatory postural reactions") would be for the language or speech in general and not the task of articulating the first phoneme/syllable. Since the subject was not presented with a list of stimuli, there was no list effect to take into account. Also, since the first word of each sentence was sufficiently varied, there was no way that the subject could predict what articulation would be necessary next. This most probably eliminated any anticipatory coarticulation effects on the ISP.

2.3.2. Data Analysis

The ultrasound data were transferred to a manipulable file format (Adobe Premiere 6.0 movie files) by means of a Sony DCR-TRV900 digital video camera connected via a FireWire cable to an Apple PowerBook G4 laptop computer. The ultrasound movie files were then cropped so that the first frame in each file was the frame immediately after the clapper was first heard. Possible periods of rest to be used for analysis were found by playing back the ultrasound movie files and searching after every sentence for a period of at least 10 frames (i.e. 333 ms) of no tongue motion in the B-mode tongue shape and the M-mode lines. A 10-frame period was chosen, as it was the longest possible rest period such that the tongue was considered to be at rest in an average of about 50% of the inter-sentential pauses across all 24 subjects. If such a period of 10 frames of no tongue motion existed, then the centre frame of that period was chosen as a "possible rest frame" for analysis ("possible" because if it was not in one of the desired phonetic contexts, it was not used). For each subject, a list of all possible rest frames was constructed for all trials that the subject completed. For each subject, Table 1 shows the total number of times after a sentence (out of 180 possible sentences, unless otherwise stated) when that subject's tongue was at a complete stop for at least 10 frames of the ultrasound movie file. Due to time constraints during data analysis, the total number of times the tongue was at rest was not investigated for 5 of the French subjects (appears as "n/a" in the table). The average number of frames used per English and French speaker was 58 and 46, respectively. The total number of rest positions analyzed was 405 from monolingual English speakers and 365 from monolingual French speakers, for a grand total of 770.

Table 1. Total rest frames available (out of 180) and number of rest frames actually used (i.e., ones found in a required phonetic context)

English Subject	Total rest frames/180	Total used	French Subject	Total rest frames/180	Total used
1	116	61	8	n/a	45
2	94	51	9	n/a	68
3	131	74	10	65	22
4	101	63	11	47	22
5	76	46	12	n/a	37
6	71	51	13	122	58
7	103	59	14	n/a	56
			15	n/a	57

Table 2. Total possible available ISPs for each pre-ISP word and each context

Context	English word	Total		French word	Total	
FrontV	Thai, July	10	**26**	ail	3	**24**
	day, holiday	11		plaie, musée, vallée	9	
	January	5		outils, nuit, radiographie	12	
BackV	Sue, through	11	**22**	perdus, trou	9	**24**
	show, scenario	11		auto, chaudron, maison	15	
Schwa	regatta	5	**5**	monsieur	3	**3**
CoronalC	again, weekend	18	**29**	assiette, recettes	6	**15**
	class	5		face	3	
	lunch	6		roche, sacoche	6	
DorsalC	spring	7	**12**	camping	3	**9**
	week	5		clinique, grecques	6	
LowV	job	9	**9**	champ, étang	9	**9**

Phonetic context was balanced by considering the IPA representation of the standard Canadian-English and Québécois-French pronunciation of the final syllable of each sentence-final word. In order to have enough tokens to do a reliable statistical analysis, it was necessary to assume that a French nasalized vowel was equivalent (in terms of the articulatory configuration of the tongue, lips, and jaw) to its non-nasalized English counterpart. Table 2 shows the final words from all of the English and French sentences for which the following rest position *was eligible* to be chosen (i.e., only if the tongue came to a complete stop for 10 frames during this time) for analysis. The rest frames were extracted from the video files and saved as .tiff image files.

In order to determine correctly which Optotrak frame corresponded to a given ultrasound frame, it was necessary to search through the Optotrak data for marker 12 (the clapper marker) and find the lowest vertical position for the marker (lowest x-coordinate in the Optotrak coordinate system) after the clapper was dropped. In some cases the clapper bounced, resulting in marker 12's vertical position increasing slightly before dropping slightly again. In this case, the first minimum was taken ("clprmin"). This frame where the marker first reached its minimum value was taken to correspond to the ultrasound video frame where the clapper noise was first heard. Because the Optotrak data was collected at 90 Hz, whereas the ultrasound data was at 29.97 Hz, a formula was used in the main MATLAB program to calculate the Optotrak frames of interest based on the ultrasound frames of interest. The ultrasound frames of interest were simply multiplied by 90, divided by 29.97, and then the result was added to "clprmin" to get the Optotrak frames of interest.

In each trial, the frame where the alveolar ridge was the most clearly visible was chosen and saved as a .tiff image file. These alveolar ridge files were later used in a MATLAB program to define the (constant) location of the alveolar ridge with respect to the four glasses markers in each trial (i.e. the coordinates of the alveolar ridge in head space). This calculation of the position of the alveolar ridge in head space was accomplished by first using ultrasound data to calculate the location of the alveolar ridge relative to the probe in ultrasound image space, then using Optotrak data to calculate the location of the ultrasound probe with respect to the head. Knowing the alveolar ridge relative to the probe, and the probe relative to the head, gave us the position of the alveolar ridge relative to the head. Then knowing from the Optotrak data how the head moved about the probe during the course of a trial, we then knew how the alveolar ridge moved about the probe and we determined the coordinates of the alveolar ridge in all ultrasound frames of interest.

As mentioned above, a series of MATLAB programs ("m-files") were used for data organization and analysis. Certain functions contained in the MATLAB Image Processing Toolbox were also used by the m-files. Optotrak data in its original floating point file format was converted into MATLAB 3D matrixes by a program supplied by Mark Tiede (Haskins Laboratories / MIT). These 3D matrixes were used by the main m-file, which was written by the present author. This main m-file, in which measurements were made and calculations were performed, had as its input the rest position .tiff images, the alveolar ridge .tiff images, and a database of Optotrak numerical values from three other m-files.

The articulator measurements that were relevant for this experiment and on which statistical analyses were performed are shown in Table 3. In this paper, these 12 measurement locations are hereafter referred to as the "ISP components".

Table 3. Definitions of the ISP components used in statistical analyses

TTht	distance from probe centre (a point exactly 1 cm below the surface of the probe on the midsagittal line and marked on the ultrasound) to tongue tip
TBht	distance from probe centre to tongue body
TDht	distance from probe centre to tongue dorsum
TRrt	distance from probe centre to tongue root
JAWl	amount of jaw lowering from maximally closed position
ULlo	upper lip height relative to glasses
LLlo	lower lip height relative to glasses
ULpr	upper lip protrusion - distance from midsagittal upper lip marker to plane constructed through alveolar ridge and two end points of glasses
LLpr	lower lip protrusion - same as upper lip, but using lower lip marker
Lvap	vertical lip aperture
Lhap	horizontal lip aperture
Lnar	amount horizontal lip aperture is narrowed from maximally spread position

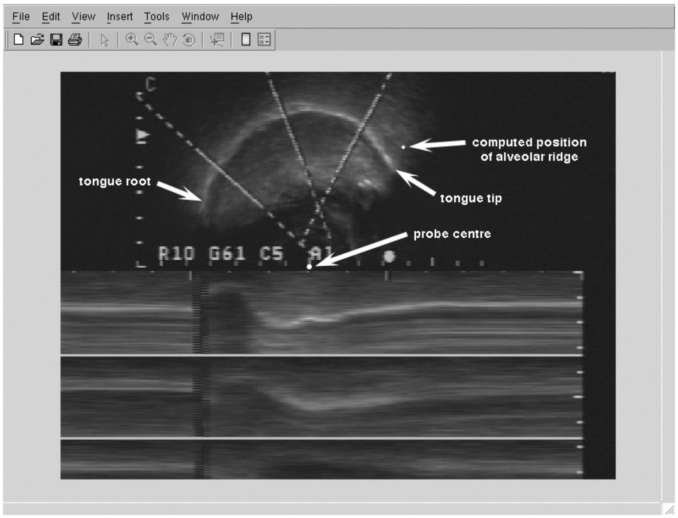
Coronal tongue shape was not measured, but it is admittedly an important factor to consider. It can be viewed with ultrasound, but with 2D ultrasound, it is not possible to see both midsagittal and coronal views of the tongue simultaneously.

The procedure that the MATLAB m-files followed was first to prompt the user for the subject's 3-letter code name and a Trial number to analyze. After retrieving relevant data for the specified trial of that specified subject, the program then displayed the stored .tiff image of the frame where the alveolar ridge was visible. The program prompted the user to click on the location of the alveolar ridge.

After the user clicked on the alveolar ridge, the program retrieved the first stored rest position .tiff image, and placed two red marks on it, one at the probe centre (10 mm below the surface of the probe) and the other at the point where the previously-clicked-on alveolar ridge was now computed to be after corrections for head movement. Note that in order to register the ultrasound images in a physical space defined by the Optotrak, a simplifying assumption was made that the ultrasound images always showed the midsagittal plane. This allowed the 3D coordinates of the alveolar ridge to be mapped onto the 2D ultrasound image by simply ignoring the third coordinate (i.e. the one off the midsagittal plane). Although it is very likely that the ultrasound images were not always showing the midsagittal plane, in a preliminary analysis of a subset of the data reported here, Gick et al. (2005, p. 512) showed that during ISP, the variation in head position in the direction perpendicular to the midsagittal plane was 1.86 mm, the smallest of the three possible translational movements.

After zooming in on the tongue, the image was then displayed to the user (e.g., see Figure 1) and the user was prompted to click on the image enough above the “hyoid shadow” that a straight line drawn to the probe centre would intersect the tongue line. The hyoid shadow is the dark area to the lower-left of the tongue root, a shadow caused by the absorption of the ultrasound waves by the hyoid bone.

Figure 1. Ultrasound frame in MATLAB of an ISP to be analyzed. The user is separately instructed to click above the hyoid shadow in the picture.



After the user clicked above the hyoid shadow, a straight line was drawn through this point and the probe centre. A second straight line was drawn through the alveolar ridge and the probe centre. Finally, two more lines were drawn that trisected the angle between the first two lines. The user was then prompted to click on the four points where each line intersected the surface of the tongue, and to do this in order from right to left (i.e. TT to TR). These measurement locations shall be called tongue tip (TT), tongue body (TB), tongue dorsum (TD), and tongue root (TR), and they correspond roughly to constrictions in the alveolar, palatal, uvular, and pharyngeal regions. Although the tongue line appears to be a thick white line, the actual surface of the tongue is the bottom edge of the white line, where it meets the black area. In the case of the tongue line not being visible, the user was prompted to click in a far corner of the image and such points were later eliminated from consideration in the analysis. After the user clicked on the four tongue points, the distance in mm from the probe centre to each point was calculated and saved.

Before any statistical analyses were performed on the data, the data was normalized. Every speaker has a different sized vocal tract, and when comparing groups of speakers across languages, normalizing the articulatory measurements is likely to reduce some of the noise in the data. Although no perfect method of normalizing speech data from different speakers has been discovered yet, a number of methods have been used in other studies (see below). The method of normalization used in this paper was to multiply each subject's data measurements by a factor that was calculated using the distance from each subject's nose bridge (as approximated by the centre glasses' marker) to the alveolar ridge (as seen in some ultrasound images). This is effectively an anatomical measure that approximately varies with some aspects of the size of the vocal tract. The multiplication factor for a given subject was the largest subject's distance (in this paper, that of Subject 6) divided by the given subject's own distance. Table 4 shows the mean distance from a given subject's nose bridge (as approximated by the centre glasses' marker) to the alveolar ridge (as seen in some ultrasound images and then calculated for each ISP). Table 4 also shows a ranking of the subjects from longest (1) to shortest (15). The mean of the mean distances for the seven English subjects was 72.19 mm, and for the eight French subjects it was 74.75 mm.

Table 4. Mean distance in mm (and standard deviation) from subject's nose bridge to calculated position of alveolar ridge; Rank (1=longest; 15=shortest)

English Subject	Distance (s.d.)	Rank	French Subject	Distance (s.d.)	Rank
1	69.11 (5.69)	13	8	75.81 (1.23)	6
2	66.44 (1.87)	14	9	79.16 (3.67)	3
3	77.85 (2.76)	4	10	63.12 (4.79)	15
4	69.55 (2.12)	11	11	76.50 (2.28)	5
5	70.95 (2.90)	10	12	80.21 (1.98)	2
6	81.90 (2.11)	1	13	74.44 (1.33)	8
7	69.52 (1.52)	12	14	75.26 (3.35)	7
			15	73.48 (1.67)	9

In Table 4, the standard deviations indicate that there is a reasonably high degree of variability in the distance from the nose bridge to the alveolar ridge. Ideally, this is a measurement that should not vary at all, assuming the glasses do not move relative to the skull. The standard deviation ranges from a low of 1.23 (Subject 8) to a high of 5.69 (Subject 1). The most probable reason for the high standard deviation in some subjects is trial-to-trial variation in the selected location of the alveolar ridge. This variation would have been due to a lack of clarity in the ultrasound frames where the subjects were swallowing. It is possible that what looked like the alveolar ridge was actually not so in some trials. If anything, this extra noise would reduce the number of significant differences found across speakers and languages, and should not introduce artificial significant effects.

Although it is intuitively apparent that tongue dimensions should vary with body size, just as across the animal kingdom, the size of the brain increases with body size (Seyfarth & Cheney, 2002), results have been mixed. Tongue measurements taken of 35 healthy Caucasian dental students by Oliver and Evans (1986) showed that the mean length, breadth, and thickness of the tongue is greater for males than for females. Note, however, that Chiang, Lee, Peng, and Lin (2003), who studied 20 Chinese medical students, found no significant difference between the 10 females' and the 10 males' mean tongue thicknesses (as measured with ultrasound from the mylohyoid muscle to the tongue body). In a three dimensional study of 25 Japanese female adults, Takada, Sakuda, Yoshida, and Kawamura (1980) showed a significant correlation between tongue volume and both the capacity of the oral cavity and the depth of the floor of the mouth, but not the height of the palatal vault. Thus, the anatomically-based method of normalization used in this paper is probably not perfect, but is probably an improvement over using non-normalized data.

3. Results

English group means are compared to French group means, in order to test the hypothesis that the English ISP is different from the French ISP. For each measurement (e.g. tongue tip height, upper lip protrusion, etc.), group means and standard deviations were calculated for English and for French. Each group mean and standard deviation is the mean and standard deviation of the individual subject means for that measurement and that language. The English and French *group* means, *between*-subject standard deviations for each language and each measurement, as well as results of *t* tests are given in Table 5.

As can be seen in Table 5, significant differences between the English and French groups were found for tongue tip height (English higher than French), upper lip protrusion (English more protruded than French), lower lip protrusion (English more protruded than French), and degree of lip narrowing - the amount that the corners of the mouth are drawn in towards the midsagittal plane from a maximally spread position (English more narrowed than French).

Table 5. Means and between-subject standard deviations (in parentheses) of English and French groups for each ISP component. Results of 12 *t* tests (assuming unequal variances). Because of the assumption of unequal variances, actual degrees of freedom are fewer than 13. See Wilson (2006) for exact degrees of freedom.

ISP Comp.	**English group mean (*SD*)**	**French group mean (*SD*)**	**Result of *t* tests**
TTht	63.88 mm (5.02)	58.35 mm (3.57)	Eng sig. higher $t(13)=2.43$; $p=.0340$ *
TBht	66.41 mm (4.55)	63.33 mm (5.46)	Eng tending higher $t(13)=1.19$; $p=.2542$
TDht	58.19 mm (8.83)	56.39 mm (5.77)	no difference $t(13)=0.46$; $p=.6560$
TRrn	48.60 mm (9.59)	48.69 mm (6.65)	no difference $t(13)=0.02$; $p=.9848$
JAWl	6.36 mm (4.16)	6.53 mm (2.72)	no difference $t(13)=0.10$; $p=.9254$
ULlo	75.39 mm (2.49)	72.38 mm (5.07)	Eng tending greater $t(13)=1.49$; $p=.1656$
LLlo	97.14 mm (3.41)	96.16 mm (6.90)	no difference $t(13)=0.36$; $p=.7291$
ULpr	31.80 mm (6.96)	23.60 mm (4.68)	Eng sig. greater $t(13)=2.64$; $p=.0242$ *
LLpr	36.00 mm (6.90)	27.01 mm (5.15)	Eng sig. greater $t(13)=2.83$; $p=.0163$ *
Lvap	22.13 mm (3.82)	23.89 mm (4.42)	no difference $t(13)=0.83$; $p=.4217$
Lhap	61.96 mm (4.21)	60.81 mm (5.62)	no difference $t(13)=0.45$; $p=.6590$
Lnar	14.11 mm (5.99)	7.38 mm (3.57)	Eng sig. greater $t(13)=2.60$; $p=.0277$ *

4. Discussion

In this research, ISP was measured in seven monolingual speakers of Canadian English and eight monolingual speakers of Québécois French in order to test the hypothesis that the ISP for Canadian English is significantly different from the ISP for Québécois French. The results presented above partially support this hypothesis. Specifically, the results in Table 5 show that the ISP for monolingual English speakers is significantly different from the ISP for monolingual French speakers in the following ways: For English speakers, the tongue tip is higher and both lips are more protruded, and the corners of the mouth are drawn farther away from a maximum spread position than for the French speakers. These significant differences match those of Gick et al. (2004) for the tongue tip height and the upper lip protrusion, but they are opposite those of Gick et al. for the lower lip protrusion. Note that the lip protrusion results are also contrary to expectation based on the non-instrumental accounts of Honikman (1964) and others. Since Gick et al. were not able to measure lip aperture with the x-ray data they used, no comparison of lip aperture or degree of spreading can be made. Gick et al. found that the tongue body was higher for English speakers. Table 5 shows that in the present study, although the English tongue body tended to be higher, there was no significant difference between the English and French speakers (p = .2542). Also, results from Gick et al. showed that the tongue root was more retracted for English speakers. The present results show absolutely no difference in tongue root position between English and French speakers (p = .9848). Finally, neither the results from Gick et al. nor the present results show any difference in jaw height between the English group and the French group.

Thus, out of the six possible comparisons that can be made between the present study and that of Gick et al., three show the same results: the same significant differences for tongue tip height and upper lip protrusion, and the same lack of significant difference for jaw height. Of the other three comparisons that do not completely agree, two were found to be significant by Gick et al. but do not differ significantly in the present results - namely, tongue body height and tongue root retraction. Although tongue body height was not found to be significantly different between English and French in this study, the tendency was for English to be higher, the same direction as the Gick et al. results. An explanation for these two differences in results between the Gick et al. study and the present one may be the more stringent statistical method employed in the present study. As mentioned in Section 1.2, the choice in Gick et al. of using each individual token as the experimental unit for statistical comparison makes it more likely that statistically significant differences will be found. The third comparison between the Gick et al. study and the present experiment that does not agree (i.e. lower lip protrusion) is found in both studies to be significantly different across languages, but in opposite directions. One reason for this may have to do with the effect of phonetic context on the position of the lower lip. While the definition of an ISP in Gick et al. was a minimum length of 3 ultrasound frames (i.e. about 100 ms), the minimum ISP length in this study is 10 frames (i.e. about 333 ms). Thus although the articulators may have appeared to be at rest in the Gick et al. study, it is possible that there was simply not enough time for the articulators to return to a rest position, especially given the fact that the subjects could already see the next sentence and could continue reading when ready.

Although they did not measure it, Gick et al. posited that the tongue dorsum could be higher for French than for English because the other three tongue measurements all indicated that French speakers' tongues have a smaller midsagittal area than English speakers' tongues. The results of the present study do not support this view - no difference was found between the English and the French tongue dorsum height. Thus it is more likely that Gick et al.'s (p. 226) other explanation is true, namely that there could be more lateral expansion of the tongue for French speakers, and that due to the fact that the tongue is a muscular hydrostat (Kier & Smith, 1985), this lateral expansion causes a reduction in the total midsagittal area of the tongue. This explanation agrees with Honikman's (1964) assertion that the English tongue tip is "tapered" whereas the French tongue tip is "untapered".

In addition to the tongue dorsum measurement, five other measurements were made in this study that were not made in the Gick et al. study: upper and lower lip height (measured as the distance from the bridge of the nose), vertical lip aperture, horizontal lip aperture, and the distance that the horizontal lip aperture differed from a maximally spread position (i.e. "degree of lip narrowing"). Neither the lip height nor the lip aperture were different across languages, but the degree of lip narrowing was significantly greater for English, meaning the lips were closer to a maximally spread position for French. As increased lip spreading naturally decreases the amount of lip protrusion, this difference is consistent with the above findings that both lips were more protruded for English speakers.

Given the higher frequency of rounded segment types in the phonemic inventory of French, it is perhaps surprising that French had a more spread-lip ISP than English did. It is customary to think of rounding as involving lip protrusion. However, "rounding" in Québécois French actually could primarily involve a decrease in vertical lip aperture and spreading the lips could cause this decrease. This type of rounding is what Heffner (1950, p. 98) referred to as "vertical lip rounding". Heffner stated that lip protrusion is "much less frequently found with vertical lip rounding" than with horizontal lip rounding. However, if vertical lip aperture were a salient component of the ISP of the lips, then we would expect to see a cross-linguistic difference in this component (i.e. "Lvap"), but we did not. Although the type frequency of rounded segments in the phonemic inventory of French is high compared to English, it is possible that the token frequency of rounded segments in French is comparable to or even lower than that of English. In that case, the results showing French having a more spread-lip ISP than English would not be surprising. Future work relating AS to token frequencies could shed light on this issue.

Another result that at first seems surprising is the fact that the English group had a higher tongue tip during ISP than the French group did. This seems surprising given the fact that coronal consonants in French have a dental place of articulation, more anterior than English coronals, which have an alveolar place of articulation. However, considering what was actually measured by TTht, at least one reasonable explanation presents itself. The measurement denoted by TTht was the distance from the centre of the ultrasound probe to the surface of the tongue and this was measured along a line that intersected the alveolar ridge. Thus, if the tongue tip were anterior to the alveolar ridge (as it is in the case of a French coronal), then TTht would actually be measuring the height of the tongue in a location posterior to the tip (i.e. the tongue blade). During ISP, if the anterior part of the tongue were in an optimal position for articulating a coronal sound (which it may or may not be), the tongue would be higher for English than for French along the line running through the alveolar ridge.

The ISP components with the greatest crosslinguistic *similarities* were the position of the tongue root (p = .9848 across the two monolingual groups) and the jaw (p = .9254 across the two monolingual groups). Note that these two articulators somewhat determine the position of some of the other articulators. Specifically, the tongue is resting on the jaw and hence jaw height will have a strong effect on tongue height. Also, because of the hydrostatic nature of the tongue, the degree of tongue root retraction can have a great effect on the height of the tongue body. Perhaps then, the jaw and tongue root are grossly positioned (and English and French have similar gross positions for these) and then the finer adjustments are made by the rest of the components of the tongue and the lips. Note that since jaw height was not significantly different across the two languages, the difference found in tongue tip height had nothing to do with the jaw.

5. Implications and Conclusions

One important implication of these results is for the field of L2 acquisition, especially pronunciation teaching and learning. In the last 50 years, the methods and status of pronunciation teaching have fluctuated greatly (see Morley, 1991, and Celce-Murcia et al., 1996, for thorough reviews), but recently there have been an increasing number of calls for the inclusion of AS in second language teaching curricula (Brown, 1995; Celce-Murcia et al., 1996; Collins & Mees, 1995, 2003; Esling, 1987; Esling & Wong, 1983; Jenkins, 1998; Jones & Evans, 1995; Kerr, 2000; Mompeán-González, 2003; Pennington, 1996; Pennington & Richards, 1986; Rich, 2003; Thornbury, 1993). These calls and the methods that are used to teach AS exist in spite of the fact that there has been no empirical evidence for language-specific ASs. Studies of LTAS have demonstrated similarities and differences in the acoustics of two different languages, but as mentioned previously, LTAS does not necessarily directly relate to AS, and this acoustic information provided by LTAS is often very difficult if not impossible to map onto articulatory parameters for L2 learners. The results of this paper have shown that AS is indeed language specific, and have shown exactly where the relevant differences in AS occur between Canadian English and Québécois French. These results, along with those of Gick et al. (2004), provide much-needed quantitative evidence to support the teaching of AS.

This research has shown that articulatory setting (AS), observed through the window of inter-speech posture (ISP) of the articulators, is significantly different between Canadian English and Québécois French, across monolingual groups. The ISP components that differ across these languages between monolingual groups are upper and lower lip protrusion, tongue tip height, and the degree to which the corners of the mouth are drawn towards the midsagittal plane from a maximally-spread position. In Canadian English, the upper and lower lips are significantly more protruded, the tongue tip is higher, and the corners of the mouth are drawn farther toward the midsagittal plane.

Acknowledgments

The research presented here was part of my dissertation research at the University of British Columbia, and it could not have been done without the kind assistance of many people: Bryan Gick, Eric Vatikiotis-Bateson, Stefka Marinova-Todd, Jason Chang, Fiona Campbell, Ritu Kumar, Doug Pulleyblank, Anne-Marie Comte, Rose-Marie Déchaine, Joe Stemberger, Fei Xu, Maciej Mizerski, Jean-François Plante, Fernanda Almeida, Alan Hannam, Elaine Orpe, Shaffiq Rahemtulla and the late John C. Wilson.

References

Brown, A. (1995). Minimal pairs: Minimal importance? *ELT Journal, 49,* 169-175.

Bruyninckx, M., Harmegnies, B., Llisterri, J., & Poch-Olivé, D. (1994). Language-induced voice quality variability in bilinguals. *Journal of Phonetics, 22,* 19-31.

Campbell, F. M. (2004). *The gestural organization of North American English /r/: A study of timing and magnitude*. M.A. Thesis, University of British Columbia.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge, U.K.: Cambridge University Press.

Chiang, Y.-C., Lee, F.-P., Peng, C.-L., & Lin, C.-T. (2003). Measurement of tongue movement during vowels production with computer-assisted B-mode and M-mode ultrasonography. *Otolaryngology - Head and Neck Surgery, 128,* 805-814.

Collins, B., & Mees, I. M. (1995). Approaches to articulatory setting in foreign-language teaching. In J. W. Lewis, (Ed.), *Studies in General and English Phonetics: Essays in Honour of Professor J. D. O'Connor* (pp. 415-424). New York: Routledge.

Collins, B., & Mees, I. M. (2003). *Practical phonetics and phonology: A resource book for students*. London: Routledge.

Esling, J. H. (1987). Methodology for voice setting awareness in language classes. *Revue de Phonétique Appliquée, 85,* 449-473.

Esling, J. H., & Wong, R. F. (1983). Voice quality settings and the teaching of pronunciation. *TESOL Quarterly, 17,* 89-95.

Gick, B., Wilson, I., Koch, K., & Cook, C. (2004). Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica, 61,* 220-233.

Gick, B., Bird, S., & Wilson, I. (2005). Techniques for field application of lingual ultrasound imaging. *Clinical Linguistics & Phonetics, 19,* 503-514.

Hamlet, S. L., & Stone, M. L. (1981). Pre-speech posturing of the mandible in relation to jaw activity during speech. *Journal of Phonetics, 9,* 425-436.

Harmegnies, B., Esling, J. H., & Delplancq, V. (1989). Quantitative study of the effects of setting changes on the LTAS. In J. P. Tubach & J. J. Mariani (Eds.), *European conference on speech communication and technology* (Vol. 2, pp. 139-142). Edinburgh: CEP Consultants.

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

Honikman, B. (1964). Articulatory settings. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, & J. L. M. Trim (Eds.), *In Honour of Daniel Jones* (pp. 73-84). London: Longman.

Jenkins, J. (1998). Which pronunciation norms and models for English as an international language? *ELT Journal, 52,* 119-126.

Jones, R. H., & Evans, S. (1995). Teaching pronunciation through voice quality. *ELT Journal, 49,* 244-251.

Kerr, J. (2000). Articulatory setting and voice production: Issues in accent modification. *Prospect, 15,* 4-15.

Kier, W. M., & Smith, K. K. (1985). Tongues, tentacles and trunks: the biomechanics of movement in muscular-hydrostats. *Zoological Journal of the Linnean Society, 83,* 307-324.

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

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge, U.K.: Cambridge University Press.

Laver, J. (2000). Phonetic evaluation of voice quality. In R. D. Kent & M. J. Ball (Eds.), *Voice quality measurement* (pp. 37-48). San Diego: Singular.

Max, L., & Onghena, P. (1999). Some issues in the statistical analysis of completely randomized and repeated measures designs for speech, language, and hearing research. *Journal of Speech, Language, and Hearing Research, 42,* 261-270.

Mompeán-González, J. A. (2003). *Pedagogical tools for teaching articulatory setting*. Poster presented at the 15th International Congress of Phonetic Sciences (ICPhS 15), Barcelona.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly, 25,* 481-520.

Munhall, K. G., Vatikiotis-Bateson, E., & Tohkura, Y. (1994). *X-ray film database for speech research, ATR technical report TR-H-116*. Kyoto: ATR Human Information Processing Research Laboratories.

Öhman, S. E. G. (1967). Peripheral Motor Commands in Labial Articulation. *Speech Transmission Laboratory - Quarterly Progress and Status Report (Royal Institute of Technology (KTH), Stockholm), 4/1967,* 30-63.

Oliver, R. G., & Evans, S. P. (1986). Tongue size, oral cavity size and speech. *The Angle Orthodontist, 56,* 234-243.

Pennington, M. C. (1996). *Phonology in English language teaching: An international approach*. London: Longman.

Pennington, M. C., & Richards, J. C. (1986). Pronunciation revisited. *TESOL Quarterly, 20,* 207-225.

Rich, S. (2003). Introducing voice-setting phonology, part 3. Manuscript downloaded on December 1, 2003 from <http://www.developingteachers.com/articles_tchtraining/voiceset3_sarn.htm>.

Schmidt, R. A., & Lee, T. D. (1999). *Motor control and learning: A behavioral emphasis* (3rd ed.). Champaign, IL: Human Kinetics.

Seyfarth, R. M., & Cheney, D. L. (2002). What are big brains for? *Proceedings of the National Academy of Sciences, 99,* 4141-4142.

Takada, K., Sakuda, M., Yoshida, K., & Kawamura, Y. (1980). Relations between tongue volume and capacity of the oral cavity proper. *Journal of Dental Research, 59,* 2026-2031.

Thornbury, S. (1993). Having a good jaw: Voice-setting phonology. *ELT Journal, 47,* 126-131.

Wilson, I. L. (2006). *Articulatory settings of French and English monolingual and bilingual speakers*. Ph.D. dissertation, University of British Columbia.