A study of high front vowels with articulatory data and acoustic simulations

Michel T.-T. Jackson^{a)} and Richard S. McGowan
CReSS LLC, 1 Seaborn Place, Lexington, Massachusetts 02420

(Received 15 January 2011; revised 14 February 2012; accepted 15 February 2012)

The purpose of this study is to test a methodology for describing the articulation of vowels. High front vowels are a test case because some theories suggest that high front vowels have little cross-linguistic variation. Acoustic studies appear to show counterexamples to these predictions, but purely acoustic studies are difficult to interpret because of the many-to-one relation between articulation and acoustics. In this study, vocal tract dimensions, including constriction degree and position, are measured from cinéradiographic and x-ray data on high front vowels from three different languages (North American English, French, and Mandarin Chinese). Statistical comparisons find several significant articulatory differences between North American English /i/ and Mandarin Chinese and French /i/. In particular, differences in constriction degree were found, but not constriction position. Articulatory synthesis is used to model the acoustic consequences of some of the significant articulatory differences, finding that the articulatory differences may have the acoustic consequences of making the latter languages' /i/ perceptually sharper by shifting the frequencies of F_2 and F_3 upwards. In addition, the vowel /y/ has specific articulations that differ from those for /i/, including a wider tongue constriction, and substantially different acoustic sensitivity functions for F_2 and F_3 . © 2012 Acoustical Society of America. [http://dx.doi.org/10.1121/1.3692246]

PACS number(s): 43.70.Bk, 43.70.Kv, 43.72.Ct, 43.72.Ja [CHS]

Pages: 3017–3035

I. INTRODUCTION

One of the goals of phonetic linguistics is to find principles that underlie the inventories of speech sounds in various languages. Among the attempts to describe or discover principles underlying cross-linguistic preferences for particular vowel phonemes are Maddieson (1984), Ladefoged and Maddieson (1996), Liljencrants and Lindblom (1972), and Wood (1979, 1986). Among the phenomena that these works have documented or attempted to explain are the cross-linguistic preferences for five-vowel inventories and the corner vowels /i/, /a/, and /u/. Theoretical studies, including Quantal Theory (Stevens, 1972, 1989) and Dispersion-Focalization Theory (Schwartz *et al.*, 1997a,b), have suggested various possible explanations for these preferences. Both Quantal Theory (QT) and Dispersion-Focalization Theory (D-FT) suggest that /i/, in particular, does not vary much from language to language either because of its acoustic, or because of its perceptual, properties. QT suggests that the acoustic properties of a vowel produced with F_2 at its maximum, such as /i/, are insensitive to small articulatory perturbations, and thus preferred. D-FT suggests that the combination of perceptual distinctiveness (dispersion) and a preference for close formant pairs (focalization; in this case, F_3 and F_4) always drives a vowel into the high front corner of the vowel space, thereby producing an /i/ in all vowel inventories modeled in Schwartz *et al.* (1997b).

Productions of /i/ do appear to be remarkably similar cross-linguistically in some studies. For example, Strange *et al.* (2007) measure formant frequencies from vowels in North

German, Parisian French, and American English using Discrete Fourier Transform and LPC spectral analysis. The formant frequencies for /i/ as produced by their speakers are very similar; for instance the mean F_2 produced by the female speakers of North German, Parisian French, and American English varies by less than 100 Hz among the languages (values based on Strange *et al.* (2007), Appendixes A, B, and C).

However, in other studies, the acoustic properties of /i/ show significant cross-linguistic variation. Jongman *et al.* (1989), using LPC-based formant extraction, find that the vowel spaces of American English and German are expanded relative to the vowel space of Greek: the F_2 of /i/ in both American English and German is higher than the F_2 of /i/ in Greek, and the F_1 of /i/ in American English and German is lower than it is in Greek. Similarly, Bradlow (1995), using formant frequencies measured from LPC spectra and digital spectrograms, shows parallel differences between American English and Spanish: the F_2 of /i/ in American English is significantly higher, and the F_1 of /i/ lower, than those of /i/ in Spanish. Fletcher and Butcher (2002), using formant frequencies measured from digital spectrograms, show that some Australian languages with 5- and 6-vowel inventories have corner vowel qualities that are substantially centralized with respect to Australian English, and in particular that F_1 of /i/ in these languages is higher than would be suggested by QT and D-FT. Al-Tamimi and Ferragne (2005), using formant frequencies measured from LPC spectra, show that two dialects of Arabic—Moroccan Arabic with /i: ə a: u u:/, and Jordanian Arabic with /i i: e: a: o: u u:/—show similar effects, with Moroccan Arabic having a slightly smaller vowel space than Jordanian Arabic, and both having a smaller vowel space than French; F_1 of /i/ and /u/ in these dialects of Arabic is higher than F_1 in French /i/ and /u/.

^{a)}Author to whom correspondence should be addressed. Electronic mail: ladmttj@ix.netcom.com.

Larger-scale cross-linguistic acoustic studies of vowel inventories such as Livijn (2000) and Gendrot and Adda-Decker (2007) give similarly varying results. Livijn (2000), using formant frequencies measured from digital spectrograms, surveys 28 languages and shows that the languages with the largest inventories (11 vowels or more) have a tendency to have greater distances between the corner vowels, and therefore probably a larger vowel space without quantal vowel qualities or hotspots. On the other hand, Gendrot and Adda-Decker (2007, Fig. 5), using formant frequencies automatically extracted using an LPC-based method, find that some languages with smaller vowel inventories, such as Mandarin Chinese and Italian, appear to be counterexamples to this trend, having more extreme formant values for /i/, and thus greater dispersion, than the larger vowel inventories of French, English, and German.

Another line of work has concentrated on trying to demonstrate how a vowel system, or more generally, the entire sound system of a language, could evolve (de Boer, 2000a,b; Oudeyer, 2005a,b). By using various games in which software agents exchange acoustic information about synthetic vowels and some limited form of feedback, a population of agents can be shown to converge to a stable state that resembles a natural vowel inventory. These simulations use acoustic-perceptual models based on Schwartz *et al.* (1997b) and articulatory-acoustic models based largely on Maeda (1989) and Cook (1989). These studies demonstrate the feasibility of generating a vowel system with only minimal assumptions. However, these studies are not successful at modeling existing vowel systems: as Oudeyer (2005b, p. 447) remarks "... we are not able to predict systems with many vowels...." The question of whether /i/ or any of the corner vowels converges to a similar vowel quality in different vowel systems is not addressed in these simulations.

There are theoretical reasons to believe that /i/-like vowels are not all acoustically or articulatorily similar. Wood (1986) suggests that perceptual pressures within specific vowel inventories, such as the need to distinguish /y/ from /i/, will lead to acoustically and articulatorily different realizations of /i/. Wood (1986) presents evidence from articulatory-acoustic modeling that what he terms a midpalatal constriction in /i/

produces a quantal /i/, with a maximal value of F_2 . However, a more anterior, or prepalatal constriction in /i/ lowers F_2 slightly and raises F_3 strongly relative to the midpalatal constriction. The net result of the prepalatal constriction is a sharper [i.e., with formants displaced upwards in frequency; see Jakobson *et al.* (1952)] prepalatal /i/, specifically, a higher average ($F_2 + F_3$) even though F_2 is somewhat lower (Wood, 1986, p. 392). Other articulatory gestures in /y/, namely larynx lowering, lip rounding, and tongue body lowering, all tend to decrease the F_2 of /y/ relative to /i/. Because larynx lowering and lip rounding have minimal effect on F_1 and tongue lowering a large effect on F_1 with the tongue in high front position, the net effect of these maneuvers is to increase the F_1 of /y/ relative to that of /i/, making the /y/ flatter (Wood, 1986, p. 394). Thus, the prepalatal /i/-/y/ sharp-flat contrast is greater than the ordinary midpalatal /i/-/y/ contrast.

Some articulatory/acoustic data are consistent with Wood's (1986) claims. For instance, Hoole (1999), using electromagnetic articulography, shows that a pellet placed on the anterior of the tongue shows tongue body lowering (and backing) in German /y/, relative to German /i/ (Hoole, 1999, Fig. 2). Hoole (1999, Fig. 3) also shows that the F_2 of both /y/ and /i/ is relatively insensitive to changes in constriction location, as measured by changes in the location of the tongue pellet in the direction parallel to the hard palate. However, it is not clear from this language-specific data whether the constriction location in the German high front vowels is the same as the constriction location in other languages; nor is it apparent whether or not the constriction in the German high front vowels is prepalatal or midpalatal. Thus, the articulatory source of the apparent cross-linguistic variation /i/ is not resolved by this kind of data.

Nor, in general, does the *acoustic* evidence for cross-linguistic variation in the production of /i/ allow the determination of what *articulatory* variation is responsible. For example, it is not possible to determine whether the difference between the higher F_2 in French and German /i/'s and the lower F_2 in Spanish and Greek /i/'s is due to tongue or lip position differences. In general, a combination of articulatory and acoustic considerations yields clearer results. For instance, Perkell *et al.* (1993) combine articulatory and acoustic measurements to demonstrate the existence of a trading relation between lip-rounding and tongue-body raising in American English /u/.

FIG. 1. Examples of reference points P1–P4 and P0 for gridline construction: P1—alveolar ridge; P2—most superior point on the hard palate; P3—superior rear pharyngeal wall; P4—inferior rear pharyngeal wall; P0—center of the circle through P1, P2, and P3. The joint (x, y)-distribution of the points estimated from all the vowels traced for the given speaker is summarized by plotting the 2- σ ellipses. (a) reference points on a token of the vowel /i/ produced by speaker L78/79 of North American English described in Munhall *et al.* (1994); (b) reference points on a token of the vowel /i/ produced by Chinese speaker OSA described in Ohnesorg and Svarný (1955); (c) reference points on a token of the vowel /i/ French speaker S3 described in Bothorel *et al.* (1986).

FIG. 2. Examples of measurement gridlines on the same vocal tracts as in Fig. 1. Round-numbered (10, 20, ... 50) gridlines are labeled. (a) Gridlines over the same token of the vowel /i/ produced by speaker L78/79 of North American English as in Fig. 1; (b) gridlines over the same token of the vowel /i/ produced by Chinese speaker OSA in Fig. 1; (c) gridlines over the same token of the vowel /i/ French speaker S3 in Fig. 1.

FIG. 3. Sample vocal tract cross-dimension functions and constrictions. The x -axis shows the normalized axial position as a fraction of the distance measured along the vocal tract midline from the most inferior gridline to the upper incisors. The y -axis shows the normalized midsagittal cross-dimension measured along the gridline, expressed as a fraction of the mean (across tokens produced by this speaker) of the maximum cross-distances in the vocal tract back cavity. The dotted vertical line (· · ·) identifies the gridline closest to the tip of the upper incisors. The dash-dotted vertical line (- · -) identifies the approximate posterior boundary of the hard palate. Gridlines identified as being “in the constriction” of the vowel by the algorithm in Jackson and McGowan (2010) are plotted with “○.” The double-headed arrow labeled “front” shows the region of the vocal tract acoustically modeled as the front tube. The double-headed arrow labeled “back” shows the region acoustically modeled as the back cavity. Note that the laryngeal vestibule does not appear in this figure because it is modeled as the space below the first gridline. (a) Vocal tract cross-dimension function and constriction for one token of /i/ produced by speaker S3 from Bothorel *et al.* (1986). (b) Vocal tract cross-dimension function and constriction for one token of /y/ produced by the same speaker.

The goal of this study is to use acoustically motivated articulatory measures of high front vowels to find whether they detect phonemic differences within language, as well as cross-linguistic variation. The lip and tongue articulation of high front vowels in North American English, French, and Mandarin Chinese are characterized by dimensions of the constrictions between the lips and between the tongue and the hard palate. The position, length, and mean cross-diameter of these constrictions are *articulatory* properties that determine the *acoustic* qualities of the high front vowels. An articulatory synthesizer described in McGowan and Wilhelms-Tricarico (2005) was used to simulate the acoustic pattern of the measured articulatory parameters.

The first step is to describe a method for characterizing the articulation of high front vowels from cinéradiographic and x-ray images. Then this method will be shown to at least be sufficient to detect articulatory differences between /i/ and /I/ within North American English. Then, the method will be used to measure articulatory parameters for /i/ in North American English, French, and Mandarin, in order to make an articulatory comparison among the /i/’s in those three languages. Finally, the method will be applied to a comparison of /i/ and /y/ in French and Mandarin Chinese. Acoustic simulations are employed to find acoustic consequences of significant articulatory differences.

II. METHODS

A. Data

Articulatory measurements were taken from cinéradiograms and a few x rays of vowels in different languages. The languages were chosen so as to vary their vowel inventories according to the presence or absence of /y/ and size. Images of the vocal tract during vowel production from 11 different speakers of three languages were analyzed. All images were enhanced and aligned using the MATLAB Image Processing Toolkit. Hard structures in the vocal tract, especially teeth, fillings, and the hard palate, were used as landmarks for alignment, as described in Jackson and McGowan (2008).

The vowels were from North American English, French, and Mandarin Chinese. In each language, images of the entire vowel inventory were analyzed in order to place measurement gridlines above the highest larynx position in all vowels. However, the analysis was limited to the high front vowels. The vowels and number of tokens in each language are summarized in Table I.

1. North American English

North American English was chosen to provide an example of a language with a large vowel inventory without /y/. Tracings of cinéradiographic images of monophthongal vowels produced by three speakers were used, yielding the inventory /i, ɪ, ɛ, æ, ɑ, ʌ, ʊ, u/. The three speakers were the male speaker from Perkell (1969), and two female speakers from the ATR X-ray Film Database (Munhall *et al.*, 1994) originally recorded by Rochette (1973, 1977). The speaker in Perkell (1969) was age 38 at the time, originally from Ontario, Canada, had resided there for more than 20 years, and had been resident in the northeastern United States since then. The two speakers from the ATR X-ray Film Database were age 20 and 25 at the time, and both natives of British Columbia, Canada.

TABLE I. Summary of languages, speakers, vowels, and number of tokens used to construct measurement gridlines in this study.

Language	Speaker and source	Number of tokens											
		i	ɪ	e	æ	y	ø	œ	ʌ/ə	ɑ	ɔ	o	u
English	Perkell	1	1	-	3	1	-	-	-	1	-	-	1
	L73/74	13	12	-	8	7	-	-	-	8	-	-	3
	L78/79	8	6	-	7	8	-	-	3	6	-	-	1
French	S1	2	-	2	2	-	2	2	1	-	2	2	2
	S2	2	-	2	2	-	2	2	1	-	2	2	2
	S3	2	-	3	2	-	2	2	1	-	2	2	2
	S4	2	-	2	2	-	2	2	1	-	2	2	2
Chinese	OSA	2	-	-	-	1	-	-	1	3	-	-	1
	OSB	1	-	-	-	1	-	-	2	-	-	2	1
	A1	4	-	-	3	-	3	-	3	5	-	2	5
	A4	2	-	-	2	-	1	-	1	1	-	1	2

Original tracings from [Perkell \(1969\)](#), provided courtesy of the author, were scanned and digitized at 300 dpi. These tracings were based on cinéradiography at 45 frames per second. The vowels were produced in nonsense words of the form /hətV/. The cinéradiographic recordings of running speech in the ATR X-ray Film Database ([Munhall et al., 1994](#)) have been digitized and are available on DVD. The original films were shot at 50 frames per second. Two female speakers for whom the laryngeal structures were visible were selected; each of the speakers was recorded in two films—the first in films 73 and 74 (speaker L73/74), the second in films 78 and 89 (L78/79). The vowels were produced in a variety of sentences; the ones used in this study to measure /i/ and /ɪ/ articulations are listed in the Appendix. A single frame from each token of the each of the vowels produced by the speakers L73/74 and L78/79 was selected and analyzed. Each frame was taken at or near the extreme of mandibular motion or a period of negligible tongue movement during the vowel.

2. French

French was chosen as an example of a language with a large vowel inventory with /y/. Tracings of cinéradiographic images from running speech produced by the four speakers in [Bothorel et al. \(1986\)](#) were analyzed (S1 and S4, male; S2 and S3, female). The original cinéradiograms were filmed at 50 frames per second. All the speakers are described as being native speakers of French without regional accents: speaker S1 was born in 1957, a native of in the Lower Rhine (Bas-Rhin) department in the Alsatian region of eastern France; speaker S2, born in 1957, a native of Upper Rhine (Haut-Rhin), also in Alsace; speaker S3, born in 1954 Freiburg im Breisgau (Germany), and having resided in various locations within France and overseas; and speaker S4, born in 1957. A native resident of Bordeaux, in southwestern France for 20 years before moving to Strasbourg. The speakers were aged 23–26 at the time of the recording.

The tracings of the oral monophthongal vowels /i, e, ε, y, ø, œ, u, o, ɔ, ɑ/ were digitally scanned at 300 dpi. The vowels were produced in a variety of phrases; the ones used in this study to measure /i/ and /y/ articulations are listed in the Appendix.

3. Mandarin Chinese

Mandarin Chinese was chosen as an example of a language with a small vowel inventory with /y/. Tracings of x-ray images of vowels produced by male speakers OSA and OSB from [Ohnesorg and Svarný \(1955\)](#) and cinéradiographic images of vowels from running speech produced by male speakers A1 and A4 from [Abramson et al. \(1962\)](#) were analyzed. According to [Ohnesorg and Svarný \(1955\)](#), speaker OSA was born in 1930, a native of Peking and a speaker of Mandarin; speaker OSB was born in 1928, a native of Tientsin who lived in Peking for some time. Tientsin is commonly considered a member of the Mandarin dialect group; its differences from Peking Mandarin (mainly in the realization of tones and retroflex consonants) are not relevant in this study. According to [Abramson et al. \(1962\)](#),

speaker A1 was in his late 40s, a native of Northern China, resident in Peking for 12 years, and had been in the United States 12 years at the time of the cinéradiographic recording. Speaker A4 was in his late 30s, a native of Peking, spent the first 20 years of his life there, and had been in the United States less than 10 years at the time of the recording.

The x-ray images of productions of /i, ə, a, y, o, u/ in isolated words¹ in [Ohnesorg and Svarný \(1955\)](#) were made with an exposure time of 0.03 s and traced by the original authors. The authors describe the images these tracings are based on as based on the “culminating phase” of the speech sound: “Les épreuves de tous les sons ont été prises dans leur phase ‘culminante’” [“The examinations of all the sounds were taken during their culminating phase”; [Ohnesorg and Svarný \(1955\)](#), p. 29]. These tracings were scanned and digitized at 300 dpi. The original 16 mm film, described in [Abramson et al. \(1962\)](#), contained frontal visual and audio recordings, and lateral cinéradiographic recordings of the utterances. The cinéradiographic recordings were shot at three times the normal rate for 16 mm film (i.e., 3 × 24 frames per second = 72 frames per second), but cinéradiographic portions of the movie are played back at 24 frames per second, therefore showing the articulations at 1/3 speed. The cinéradiographic portions of the film are dubbed with “stretched sound” to give approximately synchronized audio. The entire film, with both regular-speed video recordings and then 1/3 speed cinéradiographic recordings, was transferred to Betamax- and VHS-format videotape in earlier work. The VHS-format videotape was then digitized and transferred to DVD. Single frames at or near the extreme of mandibular motion or a period of negligible tongue movement from multiple tokens of the vowels /i, ə, a, y, o, u/ and [ɛ] (an allophone of /ə/) produced by speakers A1 and A4 read from a word list were selected and analyzed.

Since this data was gathered by a number of different experimenters in a number of different studies, from a number of different languages, the current study made no attempt to determine or classify the consonantal, syllabic, prosodic, or grammatical context of each token. Furthermore, the use of words embedded in various different phrasal or sentential positions by speakers L73/74, L78/79, S1, S2, S3, and S4 makes it impossible to control prosodic or intonational context.

All of the tokens listed in Table I were used to construct measurement grids according to the procedure described below. After constructing the measurement grids, tokens of the vowels /i/ and /y/ were analyzed. For North American English tokens of /ɪ/ were also analyzed. The words containing the vowels /i/ and /y/ used in this study are listed in the Appendix.

B. Measurement procedures

The measurement procedure used in this study was based on the method described in [Jackson and McGowan \(2008, 2010\)](#). For each speaker, a uniform number of gridlines was placed between the upper incisors and the highest position of the glottis in order to normalize the distribution of gridlines. This normalization allows measurements from different sources, some of which do not have any distance calibration, to be compared.

The first step in the procedure for constructing the measurement grid was to trace the midsagittal outlines of the vocal tract structures. More than 300 outline points on the upper and lower lip surfaces, hard palate and upper incisor outline, soft palate outline, tongue, laryngeal structures, and dorsal wall of the pharynx were selected from each of the digitized images described above. Points on the outlines of the other upper teeth, especially the molars, were also selected where available. Bicubic splines interpolating the selected points were used to connect the outline points with smooth curves that were used as the traced outline of the vocal tract image. Adopting Westbury's (1994) coordinate system, the maxillary occlusal plane (MaxOP) was used as the horizontal axis, and the tip of the upper incisors was used as the origin. The MaxOP was estimated by finding the line through the tip of the upper incisors that was tangent to the outline of the other upper teeth.

Approximately 53 measurement gridlines for each speaker were then constructed based on the methods in Jackson and McGowan (2008, 2010). As in Jackson and McGowan (2008), all the tokens produced by each speaker were traced to ensure that all of each speaker's gridlines were above the highest position of the glottis. Each measurement grid was based on four reference points, denoted P1 through P4. Figure 1 shows examples of these reference points for several speakers.

P1 was defined as a point on the alveolar ridge. On each tracing, an inflection point at the juncture between the incisor and the alveolar socket in the maxilla was identified visually and taken as the estimated location of P1. The mean position (across all the tracings of all the vowels for the speaker; see Table I) of these points for each speaker was used as reference point P1.

P2 was defined as the most superior point of the hard palate, in the coordinate system based on the MaxOP. The highest point on the bicubic spline tracing the outline of the hard palate was determined automatically for each image. Although it is possible that the hard palate could contain a small region that appeared to be horizontal, in fact none were found, and the highest point on the hard palate was unique on each tracing. The mean position of these points for each speaker was used as reference point P2.

P3 and P4 were located using other radiographic landmarks. In the data from North American English and Mandarin Chinese, the vertebrae were either traced by the original authors or imaged, and P3 was taken as the point on the rear pharyngeal wall at the same height as the most anterior point of the anterior tubercle of the atlas. In the data from French, the vertebrae were not imaged, and the point on the rear pharyngeal wall that was at the level of MaxOP was used. A location for P3 was estimated on each image for each speaker, and the mean position (within speaker) was used as the fixed location of P3 in subsequent analysis. Following Tiede (1996), the point on the rear pharyngeal wall at the same height as the point at the bottom of the vallecular sinus was adopted to be P4.

Finally, the center of the circle passing through P1, P2, and P3 is denoted P0, an auxiliary point required to construct the set of gridlines for each speaker.

1. Pharyngeal and oral cavity gridlines

A measurement grid was constructed for each speaker. The first gridlines constructed were in the lower pharynx, perpendicular to the P3-P4 line. They were constructed so that the topmost gridline passed through P0 (i.e., a radius of the circle passing through P1, P2, and P3) and the lowest gridline was above highest vocal fold position in the inventory of vowel images for the given speaker. In this way, there was no missing data on any of the lower gridlines due to the vocal folds having risen above them. These gridlines were approximately perpendicular to the axis of the pharynx and covered the region of the vocal tract from the larynx to the oropharynx. The spacing between the gridlines was adjusted so that a total of 53 gridlines were in the portion of the vocal tract between the upper incisors and the highest position of the glottis. The distance between the lowest gridline and the estimated position of the glottis was measured for each token. Figure 2 shows the gridlines for three speakers.

The second set of gridlines was constructed using the topmost of the previous gridlines as a starting point. From that gridline, which passes through P0, radial gridlines centered at P0 were constructed every 5° until the gridlines were within 5° of being perpendicular to the P1-P2 line segment, i.e., until the gridlines (nearly) coincided with the perpendicular bisector of the P1-P2 line segment. These gridlines covered the oral cavity-pharyngeal cavity junction.

From the last radial gridline, which was within 5° of being perpendicular to the P1-P2 line, gridlines perpendicular to the P1-P2 line were placed as far as P1. The spacing between these gridlines was the same as the spacing between the gridlines in the lower pharynx. These gridlines covered the vocal tract in the region of the anterior of the tongue and the anterior of the hard palate.

2. Anterior gridlines

In images or tracings in which the lips were well-imaged, it was also possible to construct anterior gridlines through the opening of the oral cavity formed by the lips. In practice, this was possible for all speakers in this study except the two in Abramson *et al.* (1962).

First, two anchor gridlines were drawn. The posterior anchor gridline was taken at the minimum distance between the upper and lower lips, which was also a point at which the tangents to the upper and lower lips were parallel and the normal vectors of the upper and lower lips coincided. The anterior anchor gridline was the osculating line connecting the anterior surface of both lips. This line connected both lips near their most anterior points and was tangent to both lips.

Additional gridlines were evenly distributed between the most anterior gridline of the original 53 at the alveolar ridge and the posterior anchor gridline so as to make the distance between the midpoints of the gridlines the same as the distance between the gridlines immediately before the alveolar ridge. The angular difference between the last tongue-to-hard-palate gridline and the posterior anchor gridline was also evenly distributed over the additional gridlines. Similar gridlines were placed between the posterior and anterior anchor gridlines. Figure 2 shows these anterior, as well as the oral and pharyngeal, gridlines for three speakers.

Since the position of the lips varied from vowel to vowel, the number of gridlines used to cover the interval from the alveolar ridge to the end of the lips differed for different vowels. It is worth noting that individual differences, such as degree of overbite, also contribute to variation in gridline placement anterior of the alveolar ridge. Speakers with relatively greater degrees of overbite often appeared to have an airway that turned abruptly from the oral cavity to the channel between the lips.

3. Vocal tract cross-distances

Midsagittal cross-distances were measured along each gridline for each vowel token from each speaker. Where there was an acoustically relevant airspace connected to the main airway, the midsagittal cross-distance of that airspace was included, but the thickness of the tissue separating the airways was excluded. For instance, the thickness of the uvula was excluded by subtracting it from the total distance between the tongue and the pharyngeal wall. The cross-distances were normalized as described below in Sec. II B 5.

4. Four vocal tract regions

One of the acoustically relevant features of the vocal tract shape in high front vowels is the constriction formed by the tongue and hard palate. This vocal tract constriction was characterized using the method described in Jackson and McGowan (2010). In this method, a criterion cross-distance for the constriction was defined to be 2.25 times the mean of four distances: the minimum cross-distance in the hard palate region, and the cross-distances at the three gridlines posterior to the gridline with the minimum cross-distance. Contiguous gridlines with cross-distances less than or equal to this criterion were taken to constitute the constriction. The portion of the vocal tract from the most inferior gridline (the highest glottal position) to the posterior end of the constriction was taken to be the back cavity, and the portion of the anterior end of the constriction to the last anterior gridline at the lips was taken to be the front tube.

The region from the glottis to the most inferior gridline in the pharynx is considered to be a part of the laryngeal vestibule. Thus the vocal tract is conceived to consist of four regions: the laryngeal vestibule, the back cavity, the constriction, and the front tube. This is appropriate when articulatory differences among high, front vowels are recast in terms of the first three formant frequencies (Stevens, 1998; Story, 2004). Figure 3 shows a pair of midsagittal cross-distance functions derived from the measurements along the gridlines. In Fig. 3, the front tube, constriction, and back cavity are identified. The laryngeal vestibule does not appear, since it is the space below the first gridline (cf. the vocal tract outlines and gridlines in Fig. 2).

5. Axial and cross-distance normalization

Axial distances within the vocal tract were normalized relative to the distance measured along the vocal tract midline from the most inferior gridline to the upper incisors. Positions within the vocal tract were expressed in normalized coordinates, with 0 being the highest position of the

glottis, negative coordinates being below the highest position of the glottis, 1.0 being the position of the upper incisors, and coordinates greater than 1.0 being anterior to the upper incisors. Comparisons between constriction position, constriction length, and cavity length in different vowels were in terms of this normalized axial distance, and not in terms of physical distance. This is an acoustically motivated scaling because ratios of formant frequencies are invariant as all sections of the vocal tract have their axial lengths scaled by the same factor.

Vocal tract cross-distances in the vowel tokens of /i/, /ɪ/, and /y/ analyzed in this study were normalized within each speaker. Acoustically, the ratio of cross-sectional areas of adjacent tube sections determines formant frequencies. The cross-sectional area, A , at an axial location can be related to the cross-distance, d , by a parametric function power law, known as the α - β model. The model is $A = \alpha d^\beta$, where the parameters, α and β , are functions of axial location and phonemic identity. From previous work (e.g., Ericsson, 2005), it is justified to assume that the parameters remain invariant with respect to phoneme within the restricted class of high front vowels under consideration here. Further, in the set of high front vowels, the variation of β with axial location is not too great. [We obtain an average $\beta = 1.16$ with a standard deviation of 0.26 over the 14 image planes for the female subject of Ericsson (2005) producing high vowels.] Thus, to a first approximation, it is the ratios between the cross-distances of adjacent tube sections that determine the formant frequencies.

In order to normalize the vocal tract cross-distances in an acoustically motivated way, the cross distances were divided by the mean, across all front high vowel tokens produced by the speaker, of each vowel token's maximum cross-distance in the vocal tract back cavity. A more sophisticated normalization would account for the variation of β with axial location. The variation of β and cross-sectional area with axial position is accounted for in the synthesis described in Sec. II B 7 below.

6. Measured articulatory parameters

The articulatory parameters of interest in the measurements considered here were motivated by the acoustics of a vocal tract sectioned into four regions. The length along the midline from the glottis to the most inferior gridline characterized the length of the laryngeal vestibule. Other than providing the length normalization for cross-distances and cross-distances for synthesis, back cavity dimensions were not considered. The measured constriction parameters were calculated from the gridlines within the constriction, as defined in Sec. II B 4. The length of the constriction was the distance from the most posterior constriction gridline to the most anterior constriction gridline. The mean cross-distance in the constriction was calculated from the cross-distances measured along these gridlines. As in Wood (1979, 1986), the position of the center of the constriction, which is known as the constriction position in this paper, was measured as a distance from the incisors. The constriction position was taken to be the midpoint of the constriction—halfway

between the anterior and posterior ends of the constriction. The measure of the cross-distance in the front tube was the mean cross-distance between the lips along three gridlines centered on the gridline constructed at the position of the minimum cross-distance between the lips. The effective length of the anterior vocal tract was the distance along the vocal tract midline from the anterior end of the constriction to the midpoint of the gridline at the minimum cross-distance between the lips. Because protrusion of the lips and lowering of the larynx both lengthen the vocal tract, these measures were added to make a single dependent variable, excess length, in the statistical analyses. This kept the number of the articulatory dependent variables that are acoustically relevant to a minimum.

Figure 3 shows two examples of midsagittal cross-distances, measured along the gridlines and plotted as a function of position in the vocal tract. Figure 3(a) shows one token of /i/, and Fig. 3(b), one token of /y/, produced by speaker S3 from Bothorel *et al.* (1986). The dotted vertical line ($\cdots$) identifies the gridline closest to the tip of the upper incisors. The dash-dotted vertical line ($- \cdot$) identifies the approximate posterior boundary of the hard palate. For the token of /i/ [Fig. 3(a)], the algorithm placed the posterior end of the constriction at approximately 0.69 in normalized coordinates (approximately gridline 36) from the highest position of the glottis and the anterior end of the constriction at 0.96 (gridline 50). The constrictions are denoted by the points plotted with 'O.' For the token of /y/ in Fig. 3(b), the algorithm placed the constriction from 0.58 to 0.96 in normalized coordinates (gridlines 31 to 50). The position of the constriction (halfway between the ends of the constriction) was therefore 0.83 in Fig. 3(a) and 0.77 in Fig. 3(b). From the anterior to the posterior end of the constriction gives an approximate constriction length [0.31 in Fig. 3(a); 0.42 in Fig. 2(b)]. The degree of the palatal constriction was measured in the high front vowels /i/ and /y/ by calculating the mean cross-distance of the vocal tract along the gridlines in the constriction. The mean cross-distance in the constriction in the token of /i/ in Fig. 3(a) was 0.19, expressed as a fraction of the speaker's mean maximum back cavity cross-distance; the mean-cross-distance in the constriction in the token of /y/ in Fig. 3(b) was 0.21.

7. Articulatory parameters for synthesis

To illustrate the effects on F_1 , F_2 , and F_3 values caused by differences in measured articulations, articulatory synthesis was used. A token of /i/ produced by a North American female speaker, L73/74, was used to provide a North American English base model /i/ from which differences among languages' /i/ production could be simulated. The following describes how the original /i/ was synthesized. As in Sec. II B 6, the length of her laryngeal vestibule was estimated to be the distance from the glottis to the most inferior grid line in the pharynx. For purposes of synthesis, a reasonable estimate of the area of the laryngeal vestibule was gleaned from the literature (Story, 2004), and this was used to set the cross-dimension of her laryngeal vestibule. Her back cavity cross-distances were unaltered from the original token of /i/

for all the syntheses. Her constriction channel was modeled so that the cross-distances form a parabolic function that matched her cross-distance data at the ends of the constriction channel. The minimum cross-distance within the constriction channel was at the same location as given in the data for the /i/. The cross-distance at that point was set so that the average cross-distance within the constriction channel matched that found in the data. The front tube cross-distances were determined by linear interpolation from the anterior end of the constriction to the location of minimum lip constriction, and then linear interpolation from that location to the most anterior cross-distance. Area functions for the synthesizer were computed using an α - β model for area functions derived from Ericsson's data for high vowels (Ericsson, 2005). For the purposes of illustration, the vocal tract length for the /i/ was set to be 14.8 cm, which in turn determined the cross-distances (in cm) and cross-sectional areas (in cm^2). This permitted us to calculate formant frequencies that corresponded to possible human values. The cross-distances and cross-sectional areas as functions of normalized axial distance are shown in Fig. 4(a). With the synthesizer thus adjusted, the North American English base model for /i/ possessed formant frequencies $F_1 = 379$ Hz, $F_2 = 2341$ Hz, and $F_3 = 3606$ Hz. After language differences in the production of /i/ were explored, a new Mandarin Chinese and French base /i/ model was constructed in order to examine the differences in /i/ and /y/ production, as described in Sec. III D below.

Perturbations from the base articulations were in terms of percentage change in mean values, for constriction location, constriction length, average constriction cross-distance, laryngeal vestibule length, minimum front tube cross-distance, and front tube length. The acoustic sensitivity function for F_1 , F_2 , and F_3 for the North American English base model of /i/ is exhibited in Figs. 4(b), 4(c), and 4(d). The acoustic sensitivity function is the difference between potential energy density and kinetic energy density divided by the sum of the potential and kinetic energy densities (Schroeder, 1967). Small reductions in cross-sectional area in a region of positive sensitivity increase the formant frequency, and in regions of negative sensitivity a reduction of cross-sectional area decreases the formant frequency.

C. Statistical methods

Each vowel token in this study was characterized by multiple measures: constriction position, constriction length, degree of constriction, minimum cross dimension of the front tube, and the excess length formed by adding the laryngeal vestibule length and the front tube length. Therefore multivariate analysis of variances (MANOVA) using the SPSS General Linear Models procedure was used to determine whether or not there were main effects or interactions. MANOVA or some other multivariate method was required to determine whether or not a combined effect on multiple dependent variables occurs. For instance, it is possible that a vowel such as /y/ could involve varying degrees of lip protrusion (i.e., front tube length), with lesser protrusion being combined with a decreased opening between the lips

FIG. 4. Token of L73/74's production of /i/ for synthesis that is used as a base for cross-language comparisons of /i/. (a) The vertical axis is in cm for cross-distance (left scale) and cm² for cross-sectional area (right scale). Dark line represents cross-distances and the lighter line represents cross-sectional area. The horizontal axis is normalized axial location. Acoustic sensitivity functions (dimensionless) for (b) F₁, (c) F₂, and (d) F₃ as a function of the normalized axial location are shown. The vertical lines in (b)–(d) denote the extent of the constriction.

(minimum front tube cross-distance) in order to produce the acoustic effects characteristic of /y/. Without a multivariate method, it is possible that the variance in lip protrusion would be great enough to make the lip protrusion in /y/ not distinct from the lip protrusion in /i/. On the other hand, with a multivariate method, the combination of lip protrusion and lip opening in /y/ would be significantly distinct from the combination of lip protrusion and lip opening in /i/. In this study, *F*-tests based on Wilks' λ were used. When the MANOVA indicated a significant effect, the Tukey HSD test was used to make *post hoc* comparisons. SPSS version 15.0 was used for statistical tests.

The first step in the statistical analysis was to demonstrate that the acoustically motivated articulatory measures described above are sufficient to detect a difference between phonemically distinct high front vowels /i/ and /ɪ/ of North American English. Second, the articulatory measures for /i/ in the three languages were compared in order to determine whether or not there were any articulatory differences in /i/ related to the presence of /y/ in the vowel inventory. Finally, the effect of vowel inventory size on the articulation of /i/ and /y/ was investigated by comparing Mandarin Chinese and French.

III. RESULTS

A. Articulatory measures

Figure 5 presents the cross-distance functions for /i/ and /ɪ/ produced by the three speakers of English analyzed here. Figure 5(a) shows the vocal tract cross-distance as a function of position in the vocal tract for the vowel tokens

analyzed from the speaker in Perkell (1969). Figures 5(b) and 5(c) show summary data for vocal tract cross-distances for the speakers of North American English, L73/74 and L78/79, from the ATR X-ray Film Database for Speech Research. In these figures, because of the greater number of tokens for each vowel, box-plots (Tukey, 1977) are used to represent the distribution of cross-distances in the various tokens of each vowel at each gridline. Table II shows the resulting measures of laryngeal vestibule, back cavity, constriction, and front tube size. The measures shown are (1) the length of the laryngeal vestibule, i.e., the distance from the glottis to the first gridline, (2) the mean back cavity cross-distance, (3) the back cavity length, (4) the mean vocal tract cross-distance in the constriction, (5) the length of the constriction, (6) the position of the constriction, (7) the minimum cross-distance in the front tube, and (8) the front tube length. Measures of the vocal tract midsagittal cross-distance [measures (2), (4), and (7)] were expressed as a fraction of the speaker's mean maximum back cavity cross-distance. Measures of the axial distance along the vocal tract [measures (1), (3), (5), (6), and (8)] were expressed as a fraction of the distance along the vocal tract midline from the highest position of the glottis to the incisors.

Figure 6 presents the vocal tract cross-distance for /i/ and /y/ as a function of position in the vocal tract produced by the four speakers of French from Bothorel *et al.* (1986) analyzed here. Table III presents the measures of laryngeal vestibule, back cavity, constriction, and front tube size. Some speaker-specific variation is observable. Speaker S1 [shown in Fig. 6(a)], for instance, appears to keep the tongue tip relatively high in the high front vowels, leading to a constriction that extends all the way to the upper incisors. However, speaker S2 [shown in Fig. 6(b)] appears to consistently depress the tongue tip in the high front vowels, thus increasing the vocal tract cross-distances a short distance posterior to the upper incisors. This leads to a local maximum in the vocal tract cross-distance function around a normalized coordinate of 0.89 (gridline 47) in these vowels, and moves the anterior end of the constriction away from the upper incisors.

Figure 7 presents the results for /i/ and /y/ of the four speakers of Chinese analyzed here. Figures 7(a) and 7(b) show the vocal tract cross-distance for the vowel tokens analyzed from the speakers in Ohnesorg and Svarný (1955). Figure 7(c) shows summary data in the form of a box-plot of the vocal tract cross-distances for speaker A1 from Abramson *et al.* (1962). Figure 7(d) shows the vocal tract cross-distances for speaker A4 from the same source. Table IV presents the measures of laryngeal vestibule, back cavity, constriction, and front tube size for these vowels.

Finally, it can be seen that the constriction degrees from some speakers have substantially greater variability than others. This is particularly true of speakers who produced more than 3 or 4 tokens of each vowel from running speech [speakers L73/74 and L78/79 of North American English and speaker A1 of Chinese; Figs. 5(b), 5(c), and 7(c), respectively]. This is hypothesized to be due to the effects of (uncontrolled) consonantal context in the production of naturalistic running speech (see the Appendix), as opposed to speech

FIG. 5. Vocal tract cross-dimension functions for the front vowels of North American English. The x - and y -axis scales are as in Fig. 2. The dotted vertical line ($\cdots$) identifies the gridline closest to the tip of the upper incisors. The dash-dotted vertical line ($- \cdot -$) identifies the approximate posterior boundary of the hard palate. (a) Vowels produced by the speaker in Perkell (1969). (b) Box plots for the front vowels produced by the speaker L73/74 from Munhall *et al.* (1994). (c) Box plots for the front vowels produced by the speaker L78/79 from Munhall *et al.* (1994). The number of tokens analyzed for each vowel is given in Table V.

examples concentrating on the demonstration of minimal contrasts (e.g., Perkell, 1969).

B. /i/ and /ɪ/ in North American English

In order to demonstrate that the articulatory measures can at least distinguish phonemically distinct vowels, a MANOVA was used to test the articulatory measures from the high front vowels /i/ and /ɪ/ of North American English. In the MANOVA, the constriction position, constriction length, mean cross-distance in the constriction, minimum cross-distance in the front tube, and the excess length (sum of the laryngeal vestibule length and the front tube length) for each token were the dependent variables; speaker and vowel were independent variables.

As shown in Table V, MANOVA found no significant multivariate interaction [at the $p < 0.01$ level; $F(10, 62) = 0.642$, n.s.], indicating that all of the speakers of North American English made parallel articulatory differences between /i/ and /ɪ/. The main effects of both speaker ($F(10, 62) = 3.637$, $p < 0.01$) and vowel ($F(5, 31) = 3.796$, $p < 0.01$) were significant.

The sources of the significant effects found in the MANOVA are also summarized in Table V. The multivariate effect of speaker was due to consistent speaker-to-speaker differences in the mean cross-distance in the constriction and the constriction length for the vowels /i/ and /ɪ/ (both significant at the $p < 0.01$ level). The multivariate effect of vowel, however, was entirely due to the difference between the mean cross-distances in the constriction of /i/ and /ɪ/ (significant at the $p < 0.01$ level). The (unweighted) mean cross-distance in /i/ was 0.27 ($\sigma = 0.10$), while in /ɪ/, the mean cross-distance was 0.39 ($\sigma = 0.13$). Thus, the mean

cross-distance in the constriction of /i/ was smaller than the mean cross-distance in the constriction of /ɪ/ reflecting the expected phonetic difference between /i/ and /ɪ/.

C. Cross-linguistic variation in /i/

We investigated whether or not cross-linguistic differences existed in the articulation of /i/ using MANOVA. Since the SPSS GLM procedure used for MANOVA normally drops observations with missing values, the two speakers of Mandarin Chinese (A1 and A4) whose lips were not imaged could not be included in a full analysis of the data. In order to find the effect of excluding the speakers whose lips were not imaged, an analysis with only constriction position, constriction length, and mean cross-distance of the constriction as dependent variables and all 11 speakers was compared with an analysis that included all the articulatory measures as dependent variables for each token but only 9 speakers. The independent variables in both analyses were language and speaker; however, since each speaker only produced /i/ in one language, speaker was nested within language (i.e., there was no main effect of speaker, and only a language-by-speaker interaction).

There appeared to be no particular difference in the pattern of results found in this comparison, although the MANOVA with no excluded speakers appeared to have slightly larger values of F .

However, none of the larger values of F crossed the $p < 0.01$ threshold for significance, and so we concentrate on the results of the MANOVA that included all the articulatory measures as dependent variables. As shown in Table VI, MANOVA found both a significant effect of language ($F(10, 40) = 3.216$, $p < 0.01$) and a significant nested effect of

TABLE II. Summary of articulatory measures from North American English speakers.

Speaker	Vowel	Larynx vestib. length	Back cavity		Constriction			Front tube	
			Mean cross-dist.	Length	Mean cross-dist.	Length	Position	Min. cross-dist.	Length
Perkell	/i/	0.030	0.69	0.61	0.18	0.37	0.82	0.47	0.10
	/ɪ/	0.017	0.62	0.51	0.42	0.47	0.76	0.51	0.10
L73/74	/i/	0.053	0.62	0.60	0.17	0.22	0.76	0.54	0.23
		0.016	0.67	0.52	0.23	0.43	0.75	0.56	0.08
		0.055	0.51	0.50	0.21	0.50	0.80	0.46	0.02
		0.040	0.57	0.57	0.19	0.37	0.80	0.49	0.08
		0.031	0.46	0.82	0.28	0.12	0.92	0.43	0.07
		0.013	0.61	0.54	0.23	0.40	0.76	0.70	0.08
		0.041	0.72	0.51	0.19	0.25	0.68	0.54	0.27
		0.036	0.74	0.51	0.23	0.46	0.77	0.52	0.07
		0.049	0.73	0.53	0.21	0.41	0.78	0.45	0.08
		0.049	0.61	0.50	0.25	0.50	0.80	0.28	0.02
		0.040	0.67	0.63	0.14	0.33	0.83	0.36	0.06
		0.041	0.72	0.53	0.26	0.41	0.77	0.66	0.10
		0.031	0.78	0.52	0.17	0.43	0.76	0.44	0.09
		/ɪ/	0.046	0.45	0.23	0.13	0.98	0.39	0.02
			0.046	0.56	0.27	0.45	0.77	0.62	0.08
			0.023	0.53	0.22	0.38	0.79	0.48	0.06
			0.047	0.69	0.37	0.55	0.77	0.57	0.04
			0.039	0.57	0.34	0.45	0.72	0.38	0.11
			0.035	0.73	0.21	0.46	0.80	0.62	0.02
			0.036	0.72	0.37	0.49	0.71	0.71	0.15
			0.040	0.52	0.29	0.44	0.77	0.65	0.08
			0.041	0.53	0.37	0.52	0.73	0.57	0.08
			0.037	0.40	0.48	0.99	0.54	0.48	0.05
			0.040	0.67	0.33	0.39	0.77	0.70	0.08
			0.032	0.60	0.26	0.47	0.80	0.36	0.02
L78/79	/i/	0.050	0.88	0.48	0.30	0.45	0.75	0.45	0.06
		0.063	0.93	0.37	0.44	0.63	0.75	0.49	0.01
		0.062	0.30	0.033	0.54	0.97	0.58	0.53	0.02
		0.054	0.83	0.38	0.35	0.62	0.74	0.56	0.03
		0.022	0.80	0.38	0.34	0.62	0.71	0.51	0.03
		0.047	0.69	0.58	0.21	0.34	0.8	0.36	0.05
		0.046	0.92	0.38	0.41	0.62	0.73	0.53	0.03
		0.060	0.82	0.45	0.42	0.55	0.79	0.57	0.01
		/ɪ/	0.046	0.46	0.66	0.97	0.56	0.52	0.01
			0.051	0.71	0.29	0.46	0.74	0.41	0.06
			0.038	0.28	0.57	0.99	0.55	0.68	0.03
			0.065	0.96	0.50	0.67	0.73	0.36	0.02
		0.058	0.83	0.37	0.44	0.63	0.75	0.56	0.01
		0.050	0.91	0.38	0.56	0.56	0.71	0.61	0.04

speaker ($F(30, 82) = 2.711, p < 0.01$), indicating both cross-linguistic and speaker-specific differences in the articulation of /i/.

The sources of these significant effects are also summarized in Table VI. The effect of language was significant at the $p < 0.01$ level on the mean cross-distance in the constriction and the minimum cross-distance of the front tube. There was also a trend ($p < 0.05$ in this data set) suggesting that the constriction length and excess length in /i/ may also be affected by language. The same pattern is seen in the nested speaker-within-language interaction effect: the nested effect is significant at the $p < 0.01$ level on the mean cross-distance in the constriction and the minimum cross-distance of the front tube in /i/; there is also a trend suggesting that the con-

striction length and excess length in /i/ may also be affected.

Because the MANOVA indicated a significant effect of language, the *post hoc* Tukey HSD test was used to identify the source of the cross-linguistic difference. According to the Tukey HSD test, the mean cross-distance in the constriction in North American English /i/ was significantly different from the mean cross-distance in both Mandarin Chinese and French /i/ at the $p < 0.01$ level. The mean cross-distance in the constriction in North American English /i/ was 0.27; the mean cross-distance in Mandarin Chinese /i/, 0.12; and the mean cross-distance in French /i/, 0.17. The mean cross-distance in Mandarin Chinese /i/ was not significantly different from the mean cross-distance in French /i/.

FIG. 6. Vocal tract cross-dimension functions for the front vowels of French. The x- and y-axis scales are as in Fig. 2. The dotted vertical line ($\cdots$) identifies the gridline closest to the tip of the upper incisors. The dash-dotted vertical line ($- \cdot -$) identifies the approximate posterior boundary of the hard palate. (a) Vowels produced by speaker S1 from Bothorel *et al.* (1986). (b) Vowels produced by speaker S2 from Bothorel *et al.* (1986). (c) Vowels produced by speaker S3 from Bothorel *et al.* (1986). (d) Vowels produced by speaker S4 from Bothorel *et al.* (1986).

Thus, the articulation of /i/ in North American English appears not to require as narrow a tongue constriction as the articulation of /i/ in Mandarin Chinese and French.

The average of the mean constriction cross-distances for French and Mandarin Chinese is 0.15, which is a 44% reduc-

tion from North American English. Synthesis with this percentage reduction in mean constriction cross-distance from the North American English base model for /i/ discussed in Sec. II B 7 reduced F_1 from 379 Hz to 258 Hz, increased F_2 from 2341 Hz to 2751 Hz, and increased F_3 from 3606 Hz to

TABLE III. Summary of articulatory measures from French speakers.

Speaker	Vowel	Larynx vestib. length	Back cavity		Constriction			Front tube	
			Mean cross-dist.	Length	Mean cross-dist.	Length	Position	Min. cross-dist.	Length
S1	/i/	0.058	0.75	0.67	0.14	0.30	0.87	0.35	0.07
		0.058	0.73	0.64	0.12	0.32	0.86	0.35	0.09
	/y/	0.064	0.71	0.62	0.15	0.35	0.86	0.06	0.06
		0.066	0.77	0.62	0.24	0.35	0.86	0.05	0.07
S2	/i/	0.072	0.66	0.59	0.14	0.36	0.84	0.26	0.08
		0.066	0.71	0.55	0.15	0.33	0.78	0.48	0.14
	/y/	0.065	0.71	0.55	0.18	0.35	0.79	0.16	0.14
		0.088	0.75	0.52	0.22	0.43	0.83	0.20	0.08
S3	/i/	0.040	0.64	0.65	0.18	0.31	0.84	0.46	0.09
		0.020	0.60	0.62	0.14	0.27	0.77	0.39	0.15
	/y/	0.038	0.55	0.55	0.20	0.41	0.79	0.10	0.09
		0.054	0.56	0.61	0.18	0.36	0.84	0.23	0.09
S4	/i/	0.048	0.70	0.64	0.20	0.26	0.82	0.31	0.18
		0.038	0.66	0.59	0.28	0.36	0.81	0.26	0.10
	/y/	0.046	0.70	0.58	0.21	0.32	0.79	0.16	0.20
		0.035	0.65	0.65	0.29	0.31	0.84	0.18	0.11

FIG. 7. Vocal tract cross-dimension functions for the front vowels of Mandarin Chinese. The x - and y -axis scales are as in Fig. 2. The dotted vertical line (· · ·) identifies the grid-line closest to the tip of the upper incisors. The dash-dotted vertical line (- · -) identifies the approximate posterior boundary of the hard palate. (a) Vowels produced by speaker OSA from Ohnesorg and Svarný (1955); (b) Vowels produced by speaker OSB from Ohnesorg and Svarný (1955); (c) Box plots for the front vowels produced by speaker A1 from Abramson *et al.* (1962); (d) Vowels produced by speaker A4 from Abramson *et al.* (1962).

3803 Hz. The directions of the formant frequency changes can be predicted from the acoustic sensitivity functions in Figs. 4(b), 4(c), and 4(d). As described in Sec. II B 7 the constriction is constructed as a parabolic function with minimum cross-distance at the location of the measured

minimum, which was about 1/3 of the constriction length from the posterior end for the North American English base model /i/, with the ends of the constriction matching the measured cross-distances at these locations. The mean cross-distance of the constriction tube was reduced here by

TABLE IV. Summary of articulatory measures from Mandarin Chinese speakers

Speaker	Vowel	Larynx vestib. length	Back cavity		Constriction			Front tube	
			Mean cross-dist.	Length	Mean cross-dist.	Length	Position	Min. cross-dist.	Length
OSA	/i/	0.033	0.78	0.60	0.11	0.15	0.71	0.91	0.28
		0.061	0.72	0.56	0.15	0.20	0.72	1.00	0.27
	/y/	0.068	0.64	0.53	0.34	0.33	0.76	0.38	0.19
OSB	/i/	0.033	0.69	0.62	0.12	0.29	0.79	0.23	0.14
	/y/	0.041	0.66	0.56	0.22	0.37	0.79	0.18	0.16
A1	/i/	0.005	0.79	0.54	0.13	0.28	0.69	—	—
		0.003	0.79	0.53	0.12	0.39	0.73	—	—
		0.006	0.69	0.58	0.16	0.31	0.74	—	—
		0.003	0.58	0.52	0.15	0.36	0.70	—	—
	/y/	0.009	0.75	0.58	0.15	0.36	0.77	—	—
		0.005	0.75	0.57	0.15	0.36	0.75	—	—
		0.003	0.57	0.53	0.15	0.36	0.71	—	—
A4	/i/	0.013	0.66	0.53	0.12	0.39	0.74	—	—
		0.014	0.71	0.53	0.12	0.44	0.77	—	—
	/y/	0.018	0.58	0.55	0.12	0.38	0.75	—	—

TABLE V. Results of two-factor MANOVA of North American English /i/ and /ɪ/ with articulatory measures as dependent variables, and Speaker and Vowel as independent variables. “Speaker* Vowel” indicates the MANOVA interaction term. The values of F and the significance level p were calculated by the SPSS GLM procedure. Where the multivariate test does not indicate $p < 0.01$, the values of F and p for the individual dependent variables are only provided for completeness; they do not indicate significance.

		Multivariate test (from Wilks' λ)	Constriction				Front tube min. cross-distance	Extra vocal tract length
			Mean cross-distance	Length	Position			
Speaker	F	3.637	19.932	8.307	3.948		0.078	2.613
	p	0.001	0.000	0.001	0.028		0.925	0.088
Vowel	F	3.796	12.671	1.393	1.339		0.433	0.340
	p	0.008	0.001	0.246	0.255		0.515	0.564
Speaker* Vowel	F	0.642	0.677	0.003	0.326		0.085	0.452
	p	0.772	0.514	0.997	0.724		0.919	0.640

decreasing the minimum cross-distance until the correct reduction in average cross-distance was obtained. Because there was no change permitted in cross-distance at the posterior and anterior ends of the constriction and the minimum cross-distance was closer to the posterior end, the effect of changes in the minimum cross-distance was greatest just posterior to the constriction. As a result, the cross-distance was reduced in regions of positive sensitivity functions for F_2 and F_3 [Figs. 4(c) and 4(d)]. A uniform reduction in cross-distance would have resulted in less dramatic, or no increases in F_2 and F_3 . With sensitivity functions of varying polarity in the constriction the details of how constriction cross-distances are changed are important.

The significant effect on minimum cross-distance of the front tube was also investigated using the *post hoc* Tukey HSD test. The minimum cross-distance of the front tubes in French /i/ (mean 0.36) and North American English /i/ (mean 0.49) were both significantly smaller (at the $p < 0.01$ level) than the minimum cross-distance of the front tube in Mandarin Chinese (mean 0.71), suggesting that the unrounded lip position in Mandarin Chinese was in fact more open than the unrounded lip position in the other languages. The average of the minimum front tube cross-distance for North American and French was 0.43, so that Mandarin Chinese had an average minimum front tube cross-distance 65% larger than this. The acoustical effect of this change was minimal: F_1 remained unchanged at 258 Hz, F_2 and F_3 increased slightly from 2751 Hz to 2753 Hz and from 3803 Hz to 3805 Hz, respectively, with increased minimum front tube cross-distance. While Fig. 4 predicts that that all three formant frequencies increase with this change, the effect was minimal because the minimum cross-distance

of the front tube was very close to the anterior of the constriction (1.02 in normalized coordinates) and the cross-distances were already relatively large in the North American English base model of /i/.

Overall, with the tighter tongue constriction, the synthesized French and Mandarin Chinese /i/ would possess a lower F_1 , and could possess greater F_2 and F_3 than the base /i/ synthesized for North American English. Thus French and Mandarin Chinese /i/ would be spectrally sharper than the North American English /i/, if the minimum cross-distance is in the posterior portion of the constriction region. Figures 6 and 7 indicate that this is usually the case for the subjects examined in this study. [It should be noted that some researchers consider diminishing F_3 – F_2 to decrease the spectral sharpness (Wood, 1986, p. 391), which is the case here for Mandarin Chinese and French compared to North American English /i/. However, increasing both F_2 and F_3 brings their average higher, as well as probably bringing F_3 closer to F_4 , which can reasonably be seen as spectral sharpening.]

The possibility of an apparent speaker-specific effect arising from different elicitation and data-collection methods was investigated by performing an additional MANOVA and three Mann-Whitney U -tests. Since multiple Mann-Whitney U -tests were performed, the Bonferroni correction for three tests was applied.

In the MANOVA, the vowel tokens from Mandarin Chinese were separated into the cinéradiographic versus x-ray groups, with the dependent variables being only the constriction position, the constriction length, and the mean cross-distance in the constriction (i.e., excluding the front tube measures). The cinéradiographic group in Mandarin Chinese consisted of the /i/ tokens from speakers A1 and A4; the

TABLE VI. Results of MANOVA of /i/ in North American English, French, and Mandarin Chinese with articulatory measures as dependent variables, and Language and Speaker as independent variables. “Language* Speaker” indicates the MANOVA nested interaction term. The values of F and the significance level p were calculated by the SPSS GLM procedure. Where the multivariate test does not indicate $p < 0.01$, the values of F and p for the individual dependent variables are only provided for completeness; they do not indicate significance.

		Multivariate test (from Wilks' λ)	Constriction				Front tube min. cross-distance	Extra vocal tract length
			Mean cross-distance	Length	Position			
Language	F	3.216	6.155	3.608	2.239		7.341	5.247
	p	0.004	0.007	0.043	0.128		0.003	0.013
Language * Speaker	F	2.711	6.594	2.808	1.426		6.847	1.743
	p	0.000	0.000	0.033	0.246		0.000	0.154

x-ray group consisted of the /i/ tokens from speakers OSA and OSB. The MANOVA did not find a significant difference between the two groups ($F(3, 5) = 11.312$, uncorrected $p = 0.011$) at the $p < 0.01$ level. However, the value of $p = 0.011$ verged on significance, suggesting the need for further investigation. Therefore, Mann-Whitney U -tests were performed on each of the constriction parameters (position, length, and mean cross-distance of the constriction) within the Mandarin Chinese data, again grouped by imaging method. The Bonferroni correction for three tests was used. The mean constriction position for the x-ray group was 0.74 ($\sigma = 0.05$), compared with 0.73 ($\sigma = 0.03$) for cinéradiographic speakers A1 and A4; the mean constriction lengths were 0.21 ($\sigma = 0.07$) and 0.36 ($\sigma = 0.06$), respectively; and the mean cross-distances in the constriction were 0.12 ($\sigma = 0.01$) and 0.13 ($\sigma = 0.02$), respectively. None of the Mann-Whitney U -tests resulted in a significant difference between cinéradiographic and x-ray groups.

Although one was not found, the possibility of a difference due to the elicitation and imaging methodology applying to two (OSA and OSB) of the four Mandarin Chinese speakers in this study cannot be disproved. The number of tokens from these two speakers is small and would not appear to have a large effect on the distribution of the measured constriction parameters for Mandarin Chinese, as shown by the MANOVA performed on the Mandarin Chinese speakers.

D. /i/ and /y/ in French and Mandarin Chinese

We investigated whether or not the articulatory measures in /i/ and /y/ differed in French and Mandarin Chinese using MANOVA. Again, two analyses were qualitatively compared—one with all 11 speakers but excluding the front tube measures with missing values; the other excluding two speakers but including all the articulatory measures. The independent variables in both analyses were language, speaker, and vowel. For both analyses, the MANOVA model included main effects of language and vowel, the nested effect of speaker (language-by-speaker), and the language-by-vowel interaction.

The results of the two analyses were qualitatively similar, with both having significant effects of language, speaker nested within language, and vowel at the $p < 0.01$ level. The language-by-vowel effect was not significant at the $p < 0.01$ level in either analysis. We will concentrate on the results of the MANOVA that included all the articulatory measures as dependent variables.

The effects found in this MANOVA are summarized in Table VII. The main effect of vowel was significant (at the $p < 0.01$) level for the minimum front tube cross-distance, mean constriction cross-distance, and mean constriction length. The main effect of language was significant (at the $p < 0.01$ level) on the constriction position, the minimum cross-distance in the front tube, and the front tube length. First the effect of vowel will be discussed and then language.

To simulate the acoustic differences between /i/ and /y/ a new model for the articulatory synthesizer was derived from the North American English base /i/ model. The mean

constriction cross-distance was reduced by 44%, as was found for Mandarin Chinese and French /i/ compared to North American English /i/ in Sec. III C. The location of the minimum front tube cross-distance was moved from 1.02 for North American English base model /i/ to the average for Mandarin and French, 1.14 (Tables III and IV). The laryngeal vestibule was lengthened from 0.015 to the average 0.03 for the same two languages (Tables III and IV). The new Mandarin Chinese and French base model for /i/ had an F_1 of 254 Hz, an F_2 of 2736 Hz, and an F_3 of 3708 Hz. Despite the fact that this new model for French and Mandarin Chinese /i/ had more excess length than the North American base /i/, it retained the higher F_2 and F_3 compared to the North American English base /i/ found in Sec. III C.

Since the articulation of /y/ involves lip-rounding, the main effect of vowel on the minimum cross-distance of the front tube is expected (Table VII). It appears that lip approximation, and not protrusion, was the most important component of lip-rounding in these tokens of /y/. The mean minimum front tube cross-distance for both French and Chinese /i/ versus /y/ was 0.45 versus 0.17, which means that the mean minimum front tube cross-distance was reduced 72% for /y/ compared to /i/. A reduction of 72% in minimum front tube cross-distance decreased F_1 from 254 Hz to 250 Hz, F_2 from 2736 Hz to 2384 Hz, and F_3 from 3708 Hz to 2888 Hz. The resulting cross-distances, cross-sectional areas, and the sensitivity functions for the first three formants are shown in Fig. 8. This configuration is a Mandarin Chinese and French model for lip-rounded /i/ (on the way to /y/), because there is one more significant articulatory adjustment before obtaining /y/ that will be discussed below. The F_2 and F_3 sensitivity functions (Figs. 8(c) and 8(d)) have changed substantially from those of the North American English base model /i/ [Figs. 4(c) and 4(d)] due to the extra excess length described above and lip rounding. A notable change is that F_2 's sensitivity is entirely positive within the constriction for Mandarin Chinese and French lip-rounded /i/ while it has both positive and negative values in the constriction for North American English base model /i/. Further the polarity of F_3 's sensitivity function has changed completely within the constriction. These patterns for the sensitivity functions in the constriction may be important, because lowering the tongue blade so that there is a more or less a uniform change in cross-sectional area will mean that F_2 can be lowered substantially without changing F_3 quite as much, thus flattening the spectrum.

Table VII indicates that the mean cross-distance in the constriction is greater for /y/ than for /i/ in both languages (French: 0.21 versus 0.17; Mandarin Chinese: 0.19 versus 0.13). The constriction length is also greater in /y/ than in /i/ in both languages (French: 0.36 versus 0.31; Mandarin Chinese: 0.36 versus 0.31). Taken together, these differences in the tongue constriction suggest that the tongue is slightly lower in /y/ than in /i/ with a 16% increase in mean constriction cross-distance. Simulating both the lip rounding and the change in mean constriction cross-distance and assuming that the increased constriction length is due to the reduced constriction degree, resulted in F_1 increasing from 250 Hz to 280 Hz, F_2 decreasing from 2384 Hz to 2226 Hz, and F_3

TABLE VII. Results of MANOVA of /i/ and /y/ in French and Mandarin Chinese with articulatory measures as dependent variables, and Language, Vowel, and Speaker as independent variables. "Language*Speaker" indicates the MANOVA nested interaction term. The values of F and the significance level p were calculated by the SPSS GLM procedure. Where the multivariate test does not indicate $p < 0.01$, the values of F and p for the individual dependent variables are only provided for completeness; they do not indicate significance.

		Constriction					
		Multivariate test (from Wilks' λ)	Mean cross-distance	Length	Position	Front tube min. cross-distance	Extra vocal tract length
Language	F	7.527	0.336	6.518	22.256	14.243	18.811
	p	0.005	0.572	0.024	0.000	0.002	0.001
Vowel	F	12.492	27.692	18.359	0.821	26.920	0.530
	p	0.001	0.000	0.001	0.381	0.000	0.479
Language * Speaker	F	4.613	3.799	2.666	4.645	7.223	5.106
	p	0.000	0.029	0.080	0.015	0.003	0.011
Language * Vowel	F	3.114	10.885	3.568	0.952	1.468	0.231
	p	0.066	0.006	0.081	0.347	0.247	0.638

increasing slightly from 2888 Hz to 2949 Hz from the Mandarin Chinese and French lip-rounded /i/ to the model /y/. The net effect of all the differences in articulation that were detected in these data is to flatten /y/ compared to /i/, with net increases in F_1 and F_3 and a net decrease in F_2 . In the simulation here, because the average cross-distance was increased by increasing the cross-distance at the point of minimum constriction, which was about 1/3 of the length of the constriction from its posterior end, where both the F_2 and F_3 had strong positive sensitivity, both formant frequencies had substantial decreases from /i/ to /y/ (Fig. 8). Lip rounding and vocal tract lengthening appear to strengthen the acoustic effect of tongue lowering, particularly for F_2 .

Language effects are now considered. Since, all other things being equal, a more posterior tongue constriction leaves a longer front tube, and vice versa, the co-occurrence of the effects on the constriction position and the excess length is not unexpected (Table VII). The mean constriction position in the Mandarin Chinese tokens of /i/ and /y/ was 0.74, slightly posterior to the mean constriction position in French, 0.82 (see Tables III and IV, respectively). As expected, the mean front tube length was conversely shorter in French (0.11) than in Mandarin Chinese (0.21), since the more posterior tongue constriction in Mandarin Chinese would result in a longer front tube.

Since front tube length was a component of the excess length, this led to a corresponding effect on the excess length (mean excess length in French: 0.16; in Mandarin Chinese: 0.25).

There was also a significant effect of language on the minimum cross-distance of the front tube (Table VII). The mean across vowel tokens of the minimum cross-distance of the front tube was 0.24 in French, versus 0.54 in Mandarin Chinese. This was consistent with the *post hoc* test results in the cross-linguistic comparison of /i/'s, which found that the minimum cross-distance of the front tube was greater in Mandarin Chinese than in French or North American English.

The effect of language on the mean cross-distance in the constriction was not significant at the $p < 0.01$ level, which was also consistent with results of the linguistic comparison of /i/'s; specifically, the *post hoc* finding that the mean

cross-distance in the constriction of /i/ in French was not significantly different from the mean cross-distance in the constriction of /i/ in Mandarin Chinese.

It can be expected that the extra excess length in the Mandarin Chinese would offset the greater degree of front tube approximation in French in terms of formant frequencies. To test this, acoustic simulations of the differences between Mandarin Chinese and French /i/ were undertaken. Starting with the combined Mandarin Chinese and French /i/ from the present section, the differences in front tube length and minimum cross-distance of the front tube between the two languages were implemented. For Mandarin Chinese /i/ $F_1 = 257$ Hz, $F_2 = 2789$ Hz and $F_3 = 3657$ Hz,

FIG. 8. Sensitivity functions after excess length and lip rounding have been added to the configuration in Fig. 3 in order to simulate a French and Mandarin "rounded /i/." (a) The vertical axis is in cm for cross-distance (left scale) and cm² for cross-sectional area (right scale). Dark line represents cross-distances and the lighter line represents cross-sectional area. The horizontal axis is normalized axial location. Acoustic sensitivity functions (dimensionless) for (b) F_1 , (c) F_2 , and (d) F_3 as a function of the normalized axial location are shown. The vertical lines in (b)–(d) denote the extent of the constriction.

and for French /i/ $F_1=257$ Hz, $F_2=2727$ Hz, and $F_3=3779$ Hz, which are fairly small differences in formant frequencies. If the excess tube length is generated by differences in constriction location rather than simply lengthening the front tube, with the French farther forward than the Mandarin Chinese, then the differences in F_1 and F_2 become more substantial. For Mandarin Chinese /i/ $F_1=269$ Hz, $F_2=2729$ Hz and $F_3=3733$ Hz, and for French /i/ $F_1=306$ Hz, $F_2=3046$ Hz, and $F_3=3743$ Hz. The differences in F_1 and F_2 can be predicted from the sensitivity function for the base North American English speaker in Fig. 4. (Care should be used because the sensitivity functions for F_2 and F_3 in the constriction region can change substantially with changes in the constriction and front tube regions, as can be seen by comparing Figs. 4 and 8.) It appears that moving the constriction forward will spectrally sharpen /i/.

Having established that the differences in excess length and minimum cross-distance of the front tube nearly offset each other acoustically, we test the effect of constriction location change on the /y/ model for combined Mandarin Chinese and French cited above. With the constriction location moved toward the posterior for Mandarin Chinese F_1 becomes 306 Hz from 280 Hz, F_2 becomes 1773 Hz from 2226 Hz, and F_3 becomes 3094 Hz from 2949 Hz. With the constriction location moved to the anterior for French F_1 becomes 295 Hz, F_2 becomes 2027 Hz, and F_3 becomes 2921 Hz. Thus F_2 is the formant frequency that changes the most for constriction location changes for /y/. The combined French and Mandarin Chinese model for /y/, which has its constriction location between that for Mandarin Chinese /y/ and French /y/, has a higher F_2 than the latter two configurations. Further, F_2 is the lower for the most posterior than for the most anterior constriction, which could help to flatten the /y/ spectrum.

Since speaker is nested within language in this design, the significant language-by-speaker interaction appears to indicate that there were also speaker-specific variations in the articulation of these vowels. Although the effect approached significance for several articulatory measures, only the minimum cross-distance in the front tube was clearly affected significantly at the $p < 0.01$ level. This appeared to be due to the fact that the per-speaker average of the minimum cross-distance in the front tube varied substantially, with the value found for speaker OSA (over 0.7; see Table IV) being an outlier compared to the values for all other speakers (in the range 0.2 to 0.3; see Tables III and IV).

IV. DISCUSSION AND CONCLUSION

The articulatory data used in this study were limited to the midsagittal plane, possessed finite spatial precision, and were usually not recorded simultaneously with the acoustic signal. Acoustic modeling helps fill the gaps for these limitations. In this work, a set of acoustically motivated articulatory measures successfully captured both phonemic and cross-linguistic distinctions in high front vowels. For instance, the data in this study showed evidence for articulatory differences among the /i/'s of three languages: the tokens of /i/ from North American English in this study

were less constricted than the tokens of /i/ from French and Mandarin Chinese. The acoustic modeling showed that these differences in constriction *degree* French and Mandarin Chinese /i/ could produce a lower F_1 , a higher F_2 , and a higher F_3 than the North American English base model of /i/. These differences would make the /i/ of the former two languages spectrally sharper than the /i/ of North American English. This result is consistent with the comparison of French and Moroccan and Jordanian Arabic made by [A1-Tamimi and Ferragne \(2005\)](#), where the two Arabic dialects do not possess phonemic /y/. As in the North American base model, F_1 in /i/ was higher in the Arabic dialects than it was in French.

On the other hand, the articulatory data did not show a difference in constriction *position* among the three languages considered. The acoustic modeling study showed that the sensitivity functions for F_2 and F_3 change polarity in the constriction region for the North American model /i/, leading to the possibility that a change in constriction position could have either perceptual sharpening or flattening effects. These polarity changes would make the acoustic changes in F_2 and F_3 due to a change in constriction position depend on the details of the constriction change.

These measures also captured a difference in the tongue positions of /i/ and /y/, suggesting that the high front rounded vowel /y/ differs from /i/ in both lip-rounding and in the degree of constriction produced by tongue. The tokens of /y/ from both French and Mandarin Chinese in this study showed a pattern consistent with a lower tongue position in /y/ than in /i/, together with the expected lip rounding. This articulatory data is consistent with observations in other languages, such as [Hoole's \(1999\)](#) data from German. Taken together, these articulatory differences spectrally flatten /y/ compared to /i/. Here, acoustic modeling showed that lip rounding caused the polarity of the sensitivity function for F_2 to be positive throughout the constriction region. Thus, increasing the average constriction cross-distance unambiguously increases F_1 and lowers F_2 . However, the change in F_3 depends on details of the constriction change.

French and Mandarin Chinese exhibited a difference in constriction location when the data for /i/ and /y/ were taken together. One way to change the location of the tongue constriction is to raise or lower the tongue tip. In the data considered in this study, the tongue tip in /i/ was generally above the lower incisors, and thus approximated to the upper incisors or to the alveolar ridge. Given this tongue position, the anterior end of the constriction extends all the way to the upper incisors. Figure 9 shows vocal tract profiles, either newly traced for this study, or re-drawn from the original sources, with an elevated anterior tongue position in /i/. Figure 10 shows some tokens in which the anterior tongue appears to be lowered in a high front vowel. This tongue positioning moves the anterior end of the constriction away from the upper incisors. In these figures, the distinction between anterior tongue elevation and depression is most easily observed by noting the position of the tongue tip in relation to the lower incisors. In this study, it appeared that having the anterior tongue elevated is more common; only a few speakers depressed the anterior tongue in high front

FIG. 9. Front vowel tokens that appear to have elevated tongue tip/blade. (a) A token of /i/ from North American English produced by the speaker in Perkell (1969). (b) A token of /i/ from French produced by speaker S1 in Bothorel *et al.* (1986). (c) A token of /i/ from Chinese produced by speaker A4 from Abramson *et al.* (1962). (d) A token of /i/ from Chinese produced by speaker A1 from Abramson *et al.* (1962).

vowels. Without further data, it is not possible to tell whether this is contextual variation, speaker-to-speaker variation or a genuine cross-linguistic difference.

Wood (1986) also discusses this kind of difference between the tongue constriction positions of /i/ or /y/, where he also employs acoustic modeling study to argue that this difference contributes to the acoustic stability of /y/ as predicted by QT. Wood (1986) raises the possibility that some speakers of languages with inventories containing /y/ may use tongue tip/blade elevation to distinguish /i/ from /y/ production. It appears in Wood (1986) that elevation of the anterior part of the tongue in this manner may result in a more anterior constriction, described by Wood as a prepa-

FIG. 10. Front vowel tokens that appear to have depressed tongue tip/blade. (a) A token of /i/ from North American English produced by speaker L78 from Munhall *et al.* (1994). (b) Another token of /i/ from the same speaker. (c) A token of /i/ from Chinese produced by OSA from Ohnesorg and Svamý (1955). (d) A token of /y/ from Chinese produced by the same speaker.

tal constriction. In the present study, there was no evidence for language-specific use of a more anterior, prepalatal constriction based on whether the language contained /y/ in its phonemic inventory because there was no effect of language on constriction position in the comparison of /i/'s from North American English, French, and Mandarin Chinese.

The present analysis does support Wood's idea that there are articulatory gestures beyond lip rounding that enhance the spectral flatness of /y/. Here it is in the form of reducing the degree of tongue constriction, which he observed from x-ray images but left unquantified (Wood, 1982, 1986). In fact, it is the lip rounding that makes reducing tongue constriction an effective way to spectrally flatten /y/. Parenthetically, we observed that making the lip constriction tighter than the one used to simulate the combined French and Mandarin Chinese /y/ may tend to decrease the effectiveness of reducing the tongue constriction in flattening /y/. However, this requires more investigation in order to be more conclusive.

The result here that French and Mandarin Chinese /i/'s are statistically indistinguishable from one another supports one of the results of Gendrot and Adda-Decker (2007, p. 1420) regarding Mandarin Chinese: that there is not clear evidence for an effect of inventory size on the global acoustic space between corner vowels. Rather, it seems to be the case that the smaller acoustic space for the vowel inventory observed in some languages is optional: languages with large vowel inventories, such as French, German, and English, may require large vowel spaces to accommodate their inventories, but languages with small vowel inventories do not necessarily have smaller articulatory or acoustic vowel spaces. On the other hand, when the rounded vowel /y/ is near /i/ in the vowel space, it is plausible that the acoustic contrast between these vowels is enhanced by making /i/ acoustically more extreme with higher F_2 and F_3 , and lower F_1 . In the data examined here this was accomplished with a tighter tongue constriction in /i/ production for the languages with /y/ than the language without /y/.

Given the limited number of subjects and languages in this study, it is clear that these conclusions cannot be generalized. In particular we cannot address the validity of either QT or D-FT. The difficulty of obtaining articulatory data sufficient for determining the acoustically significant constriction in vowels and other speech sounds is one factor limiting the feasibility of extending these conclusions. The difficulty of making cross-speaker and cross-linguistic comparisons is only partially alleviated by making strictly comparable measures according to a uniform measurement scheme. Despite the fact that much speech MRI data is still taken from static vocal tracts, rather than running speech, the increased availability of MRI data from speech improves the prospects for answering the kind of questions investigated in this study.

Two points on the relation between articulation and acoustics have arisen in this work. The first point is the need to use acoustic sensitivity functions that correspond very closely to the articulations under consideration in order to predict acoustics correctly. For instance, making a tighter constriction in the anterior part of a constriction region for the base North American English /i/ in Fig. 4(c) will lower F_2 , while for the added

excess length and lip-rounded articulation in Fig. 8(c), this gesture will increase F_2 . As an aside, caution is needed with generalities such as “tongue fronting always raises F_2 .” The second point is that spectral or feature enhancement (Stevens and Keyser, 2010) can be synergistic. For instance, extra vocal tract length and lip rounding appear to amplify the effect of reducing the constriction degree in lip-rounded /i/ production beyond what it would be otherwise (Figs. 4 and 8).

Taking articulatory measures and using articulatory synthesis to investigate the acoustic results clarifies both articulatory and acoustic details. The fundamental ambiguity of the mapping from acoustic output to articulatory configuration makes it essential to obtain both articulatory and acoustic data in order to be able to make inferences about the principles underlying phonetic linguistics.

ACKNOWLEDGMENTS

This work was supported by grant NIDCD-001247 to CReSS LLC.

APPENDIX: LIST OF UTTERANCES WITH VOWELS ANALYZED IN THIS STUDY

The underlined high front vowels in the utterances below were analyzed in this study. Note that these and other vowel tokens (see Table I) were used to construct the measurement grids.

A. North American English

Munhall, *et al.* (1994, pp. 15–16), speaker L73/74:

Film 73:	She's just <u>be</u> ing coy.
Mimi has a toy bear.	Ray likes going <u>fi</u> shing.
She <u>pu</u> lled with all her might.	It's made of <u>ti</u> nfoil.
All boys and girls like <u>ca</u> ndy.	Thy will <u>be</u> done.
Today <u>i</u> s bright and sunny.	Thou shalt not <u>ki</u> ll.
Not all Jews are Zionists.	The grass was <u>sti</u> ll moist.
They were lying in the sun.	There <u>he</u> lay <u>aslee</u> p.
Some say <u>he</u> was power mad.	It was you all along.
He's a great guy.	Fay wants to go shopping.
It's down by the <u>sea</u> .	

Munhall, *et al.* (1994, pp. 16–17), speaker L78/79:

Film 78:	Film 79:
He's such a romantic.	The <u>re</u> sult was galvanic.
That would <u>be</u> sadistic.	They portrayed <u>hi</u> m as a <u>thi</u> ef.
His name was Cornelius.	Theta <u>i</u> s a Greek letter.
Bacteria cause disease.	It was a vote of mass.
Drink to me only with thine eyes.	
Does technology <u>mean</u> progress?	

B. French

Bothorel, *et al.* (1986), speaker S1:

Mets tes beaux habi <u>t</u> s	Il <u>f</u> ume son tabac
Ma chemise est roussi <u>e</u>	Une r <u>é</u> ponse ambig <u>u</u> <u>e</u>

Bothorel, *et al.* (1986), speaker S2:

Mets tes beaux habi <u>t</u> s	Une r <u>é</u> ponse ambig <u>u</u> <u>e</u>
Donne un petit coup	Une p <u>â</u> te à choux

Bothorel, *et al.* (1986), speaker S3:

Mets tes beaux habi <u>t</u> s	Une r <u>é</u> ponse ambig <u>u</u> <u>e</u>
Louis pense à ça	Une p <u>â</u> te à choux

Bothorel, *et al.* (1986), speaker S4:

Mets tes beaux habits
Une réponse ambigue
Il fume son tabac

C. Mandarin Chinese

Ohnesorg and Svarný (1955, Figs. 27–34), speaker OSA (Peking):

/p³/ (2 tokens) “compare (v); contrast (v)”
/t^hɿ³/ “take (v.)”

Ohnesorg and Svarný (1955, Figs. 53–60), speaker OSB (Tientsin):

/m³/ “hulled uncooked rice” /n³/ “woman”

Abramson *et al.* (1962), speaker A1:

/i ⁴ / “loyalty; right conduct”	/t ⁴ / “ground”
/y ⁴ / “jade”	/k ⁴ / [tci ⁴] “tie”
/ci ⁴ / “fine”	/ky ⁴ / [tcy ⁴] “saw”
/cy ⁴ / “preface; series”	

Abramson *et al.* (1962), speaker A4:

/t⁴/ “ground”
/k⁴/ [tci⁴] “tie”
/ky⁴/ [tcy⁴] “saw”

¹It should be noted that the vowel often phonemically transcribed as /o/ is realized in simple contexts as [ɤ] [see Lee and Zee (2003)]. Two tokens of /o/ were used in this study as part of the sets of vowels used to construct measurement grids for each speaker. For speaker OSB, the word containing this vowel was transcribed narrowly with a diphthong [p^uo]; the authors remark “La deuxième phase ... pourrait être classifiée comme celle d'un o fermé. (La langue est moins élevée que pour l'u, la labialisation est plus faible.)” [“The second phase ... may be classified as a close o. (The tongue is less raised than in u, labialization is weaker.)”]; Ohnesorg and Svarný (1955), p. 49 and Fig. 60]. For speaker A1, the word containing this vowel was analyzed as “eu.” The component “e (E)” is described as “E mid (neutral) tongue position,” and the component “u (W)” is described as “W labial” [Abramson *et al.* (1962), p. 12]. Phonetically, the token of this word filmed and recorded at normal speed contains [o].

Abramson, A. S., Martin, S., Schlaeger, R., and Zeichner, D. (1962). *Mandarin Chinese X-Ray Film in Slow Motion with Stretched Sound* (Columbia University, Columbia-Presbyterian Medical Center and Haskins Laboratories, New York), pp. 1–17.

Al-Tamimi, J., and Ferragne, E. (2005). “Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French,” in *Proceedings 9th EUROSPEECH*, pp. 2465–2468.

Bothorel, A., Simon, P., Wioland, F., and Zerling, J.-P. (1986). *Cinéradiographie des Voyelles et Consonnes du Français (Cineradiography of French Vowels and Consonants)* (Institut de Phonétique de Strasbourg, Strasbourg), pp. 1–296.

Bradlow, A. (1995). “A comparative acoustic study of English and Spanish vowels,” *J. Acoust. Soc. Am.* **95**(3), 1916–1924.

de Boer, B. (2000a). “Emergence of vowel systems through self-organization,” *AI Commun.* **13**, 27–39.

de Boer, B. (2000b). “Self-organization in vowel systems,” *J. Phonetics* **28**, 441–465.

- Cook, P. R. (1989). "Synthesis of the singing voice using a physically parameterized model of the human vocal tract," in *Proceedings of the International Computer Music Conference* (pp. 69–72), Columbus, OH.
- Ericsson, C. (2005). *Articulatory-Acoustic Relationships in Swedish Vowel Sounds*, Ph.D. dissertation, Stockholm University, Stockholm, pp. 1–194, available at <http://www2.ling.su.se/staff/ericsson/thesis/index.html> (Last viewed November 9, 2011).
- Fletcher, J., and Butcher, A. (2002). "Vowel dispersion in two northern Australian Languages: Bininj Gun-wok and Dalabon," in *Proceedings of the IXth Australian International Conference on Speech Science and Technology*, edited by C. Bow and P. Blamey (Australian Speech Science and Technology Association, Melbourne, Australia), pp. 343–348.
- Gendrot, C., and Adda-Decker, M. (2007). "Impact of duration and vowel inventory size on formant values of oral vowels: An automated formant analysis from eight languages," in *Proceedings of the 16th International Congress of Phonetic Sciences* (Institute of Phonetics, Saarland University, Saarbrücken, Germany), pp. 1417–1420.
- Hoole, P. (1999). "Articulatory-acoustic relations in German vowels," in *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, 1–7 August 1999 (University of California, Berkeley, CA), pp. 2153–2156.
- Jackson, M. T.-T., and McGowan, R. S. (2008). "Predicting midsagittal pharyngeal dimensions from measures of anterior tongue position in Swedish vowels: Statistical considerations," *J. Acoust. Soc. Am.* **123**, 336–346.
- Jackson, M. T.-T., and McGowan, R. S. (2010). "A method for finding constrictions in high front vowels," *J. Acoust. Soc. Am.* **127**, EL6–EL12.
- Jakobson, R., Fant, G., and Halle, M. (1952). *Preliminaries to Speech Analysis* (MIT, Cambridge, MA), pp. 31–32.
- Jongman, A., Fourakis, M., and Sereno, J. A. (1989). "The acoustic vowel space of Modern Greek and German," *Language Speech* **32**, 221–248.
- Ladefoged, P., and Maddieson, I. (1996). *The Sounds of the World's Languages* (Blackwell Publishing, Oxford, UK), pp. 1–425.
- Lee, W.-S., and Zee, E. (2003). "Standard Chinese (Beijing)," *J. Int. Phonetic Assoc.* **33**(1), 109–112.
- Liljencrants, L., and Lindblom, B. (1972). "Numerical simulation of vowel quality systems: The role of perceptual contrast," *Language* **48**, 839–862.
- Livijn, P. (2000). "Acoustic distribution of vowels in differently sized inventories - hot spots or adaptive dispersion?," *Proceedings of the XIIIth Swedish Phonetics Conference (FONETIK 2000)*, Skövde, Sweden, 24–26 May 2000, pp. 93–96.
- Maddieson, I. (1984). *Patterns of Sounds. Cambridge Studies in Speech Science and Communication* (Cambridge University Press, Cambridge, UK), pp. 1–422.
- Maeda, S. (1989). "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic Publishers, Dordrecht, Netherlands), pp. 131–149.
- McGowan, R. S., and Wilhelms-Tricarico, R. (2005). "An educational articulatory synthesizer, EASY," *J. Acoust. Soc. Am.* **117**, 2543.
- Munhall, K. G., Vatikiotis-Bateson, E., and Tohkura, Y. (1994). *X-ray Film Database for Speech Research* (ATR Human Information Processing Research Laboratories, Kyoto, Japan), pp. 1–33.
- Ohnesorg, K., and Svarný, O. (1955). *Etudes Expérimentales des Articulations Chinoises. (Experimental Studies on Chinese Articulations)* (Czech Academy, Prague), **65**(5), 1–75.
- Oudeyer, P.-Y. (2005a). "How phonological structures can be culturally selected for learnability," *Adapt. Behav.* **13**, 269–280.
- Oudeyer, P.-Y. (2005b). "The self-organization of speech sounds," *J. Theor. Biol.* **233**, 435–449.
- Perkell, J. S. (1969). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study* (MIT Press, Cambridge, MA), pp. 1–100.
- Perkell, J. S., Matthies, M. L., Svirsky, M. A., and Jordan, M. I. (1993). "Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: a pilot 'motor equivalence' study," *J. Acoust. Soc. Am.* **93**, 2948–2961.
- Rochette, C. (1973). *Les Groupes de Consonnes en Français (Consonant Groups in French)* (Laval University Press, Quebec City, Quebec), 1–560.
- Rochette, C. (1977). "Radiologie et phonétique" ("Radiology and phonetics"), *Vie Médicale au Canada Français (Medical Life in French Canada)* **6**, 55–67.
- Schroeder, M. R. (1967). "Determination of the geometry of the human vocal tract by acoustic measurements," *J. Acoust. Soc. Am.* **41**, 1002–1010.
- Schwartz, J.-L., Boë, L.-J., Vallée, N., and Abry, C. (1997a). "Major trends in vowel system inventories," *J. Phonetics* **25**, 233–253.
- Schwartz, J.-L., Boë, L.-J., Vallée, N., and Abry, C. (1997b). "The dispersion-focalization theory of vowel systems," *J. Phonetics* **25**, 255–286.
- Stevens, K. N. (1972). "The quantal nature of speech: Evidence from articulatory-acoustic data," in *Human Communication: A Unified View*, edited by E. E. Davis, Jr. and P. B. Denes (McGraw-Hill, New York), pp. 51–66.
- Stevens, K. N. (1989). "On the quantal nature of speech," *J. Phonetics* **17**, 63–70.
- Stevens, K. N. (1998). *Acoustic Phonetics* (MIT Press, Cambridge, MA), pp. 277–282.
- Stevens, K. N., and Keyser, S. J. (2010). "Quantal theory, enhancement and overlap," *J. Phonetics* **38**, 10–19.
- Story, B. H. (2004). "Vowel acoustics for speaking and singing," *Acta Acust. United Acust.* **90**, 629–640.
- Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., and Nishi, K. (2007). "Acoustic variability within and across German, French, and American English vowels: Phonetic context effects," *J. Acoust. Soc. Am.* **122**(2), 1111–1129.
- Tiede, M. K. (1996). "An MRI-based study of pharyngeal volume contrasts in Akan and English," *J. Phonetics* **24**, 399–421.
- Tukey, J. W. (1977). *Exploratory Data Analysis* (Addison-Wesley, Reading, MA), pp. 39–43.
- Westbury, J. R. (1994). "On coordinate systems and the representation of articulatory movements," *J. Acoust. Soc. Am.* **95**, 2271–2273.
- Wood, S. (1979). "A radiographic analysis of constriction locations for vowels," *J. Phonetics* **7**, 25–43.
- Wood, S. (1982). "X-ray and model studies of vowel articulation," *Univ. Lund Phonetics Lab. Working Pap.* **23**, 1–49.
- Wood, S. (1986). "The acoustical significance of tongue, lip, and larynx maneuvers in rounded palatal vowels," *J. Acoust. Soc. Am.* **80**(2), 391–401.