PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Volumetric image-based comparison of male and female vocal tract shapes

Story, Brad, Hoffman, Eric, Titze, Ingo

Brad H. Story, Eric A. Hoffman, Ingo R. Titze, "Volumetric image-based comparison of male and female vocal tract shapes," Proc. SPIE 3033, Medical Imaging 1997: Physiology and Function from Multidimensional Images, (9 May 1997); doi: 10.1117/12.274050

SPIE.

Event: Medical Imaging 1997, 1997, Newport Beach, CA, United States

Volumetric Image-Based Comparison of Male and Female Vocal Tract Shapes

Brad H. Story[a], Eric A. Hoffman[b], and Ingo R. Titze[a,c]

[a]WJ Gould Voice Research Center
Denver Center for the Performing Arts
Denver, CO 80204

[b]Division of Physiologic Imaging, Department of Radiology
University of Iowa College of Medicine
Iowa City, IA 52242 USA

[c]Department of Speech Pathology and Audiology
National Center for Voice and Speech
University of Iowa, Iowa City, IA 52242 USA

ABSTRACT

A collection of 3-D vocal tract shapes corresponding to vowels and consonants of American English have been acquired for a 27 year old adult female subject using magnetic resonance imaging (MRI). Each 3-D shape was condensed into a set of cross-sectional areas of oblique sections perpendicular to the centerline of the vocal tract's long axis. Such a collection of areas is typically called an "area function. This set of images and subsequent area functions for the female subject compliments a previous similar study concerning an adult male subject [Story, Titze, and Hoffman, J. Acoust Soc. Am, 100(1), 1996]. It is the purpose of this paper to explore the morphological differences between the male and female subjects for three "cardinal" vowels /i/,/a/, and /u/. Comparisons have been made of the 3-D vocal tract shapes, area functions, and acoustic characteristics of the three vowels. The primary difference between genders is that the female pharynx is approximately 37 percent shorter than the male. Limited acoustic modeling has suggested that this shortened pharynx may play a significant role in defining male versus female voice quality.

Key words: MRI, speech, synthesis, vocal tract

1. INTRODUCTION

In the perception of human speech, most listeners typically have little difficulty discerning the gender of an adult speaker. However, computers tend to have great difficulty both in recognizing gender[1] as well as synthesizing a natural sounding female voice[2]. Thus, it is of interest to investigate the properties of the female speech production system and compare them to the characteristics of male speech. There are essentially two components of the speech production system that can contain the information that is unique to males and females; the voice source and the vocal tract. The voice source is comprised of the vibrating vocal folds, in which gender unique features are likely due to the lengths, masses, and elastic properties of the male or female vocal fold tissue. For example, a female typically has shorter, less massive vocal folds than an average male[3]. This leads to, among other things, a generally higher fundamental frequency of vibration than would have a male.

The second component, the vocal tract, filters the "raw" sound generated at the vocal folds, producing accentuated harmonics of the voice source in patterns such that the characteristic sounds of vowels and consonants are produced. In general, the female vocal tract has been observed to be about 15-20 percent shorter than that of a male[4,5,6] which in itself will raise all vocal tract resonances to higher frequencies. However, there has been some debate as to whether a particular male vocal tract shape is a uniformly or nonuniformly "stretched" version of the corresponding female vocal tract; "nonuniformly" implying that different regions of the vocal tract are stretched by differing amounts. Yang and Kasuya[5]. who have performed an MRI study of a male, female, and child vocal tract, have concluded that for purposes of extracting the phonetic content of speech, a uniform scaling of the female vocal tract with respect to male is adequate. Their conclusions don't, however, address the issue of gender unique voice quality or those characteristics of the voice that provide cues to the gender of a speaker. Recently, Wong et al.[7] attempted to transform the recorded speech of a male speaker to that of a female using the uniform vocal tract scaling proposed by Yang and Kasuya[5]. While the intelligibility remained high, the gender quality was confusing for listeners. Coleman[8] found that an artificially raised voice fundamental frequency, when used to excite a male vocal tract, still provided the perceptual cues to a male quality. Agren and Sundberg[9] determined that for alto (low range female) and tenor (high range male) singing voices, the main difference, with regard to the vocal tract, was that the third and fourth formant (F3 and F4) were more widely spaced for the female than for the male. They hypothesized that this characteristic could be a defining quality of male versus female since a narrow frequency spacing of F3 and F4 could accentuate two consecutive voice harmonics, thus placing them within the same critical band of the auditory system. Two, high-amplitude voice harmonics within the same critical band tends to produce a perception or roughness or hoarseness in the voice, more characteristic of a male[10]. Thus, a wide spacing of F3 and F4 may ensure that two consecutive harmonics will not be accentuated within a critical band, therefore maintaining a female quality. It has also been shown by Ingemann[11] and Schwartz[12] that listeners can detect gender from isolated fricative consonants in which there is no voice source, leaving the vocal tract as the primary element containing some gender characteristics. These studies suggest that the female voice quality is not entirely dependent on the voice source but also includes characteristics of the vocal tract. It seems essential, then, to acquire morphological data regarding the female vocal tract shape during production of speech sounds.

The use of MRI has, in recent years, become a viable option for obtaining 3-D information about the human vocal tract shape for static vowels and consonants. Baer et al.[13] demonstrated the use of MRI to directly measure the vocal tract shape for four vowels (2 male subjects). This study was the first demonstration of 3-D reconstructions of the vocal tract shape using imaging techniques. Dang et al.[14] used MRI to produce 3-D reconstructions of the nasal tract passages and sinus cavities. The nasal tract morphology was subsequently used to model the acoustic characteristics of the nasal system. MR volume imaging of the fricative consonants for male and female subjects has been recently reported by Narayanan et al.[15] and Narayanan[16]. These studies provide the most accurate information to date of the constrictions and air channels that produce the turbulence generated sound characteristic of fricative consonants. Yang and Kasuya[5], which was mentioned above, have reported area functions of five Japanese vowels acquired from a man, woman, and child. Story , Titze, and Hoffman[17] reported MRI based 3-D vocal tract shapes and area functions for 18 vowels and consonants of one male subject. These area functions were subsequently used to model static vowel sounds which were compared the natural vowel sounds of the imaged subject.

The aim of this study was to use MRI to acquire a set of 3-D vocal tract shapes for an adult female that correspond to vowels and consonants of American English. Such a set compliments a previous study concerning an adult male subject[17]. Some basic morphological and consequent acoustical differences between the male and female subjects are explored for the three vowels /i/, /a/, and /u/. The interest is to find those vocal tract differences that bear upon a male-like versus female-like quality of speech. The issue of normalizing female speech to an equivalent male for recognition of linguistic content is not addressed. It is recognized that these comparisons strictly apply to only one male and one female selected from the general population. Thus, no statistical power can obtained from the set, however, general characteristics from each subject are thought to be useful. The process of acquiring the imaging sets from one subject for many vowels and consonants (22) is a physically demanding and time consuming process. The acquisition and analysis of enough subjects to provide a statistically adequate sample would be prohibitively expensive and would take many years to complete. However, as more researchers begin to image the vocal tract the pool of data will eventually grow so that a reasonable statistical sample may be achieved.

2. IMAGE ACQUISITION AND ANALYSIS

2.1. Scanning Parameters and Protocol

Volumetric imaging of the vocal tract was performed using MRI for 22 different phoneme configurations of both a male and female subject. The vocal tract shapes for the male subject have been reported in Story et al.[17] and will be included in this paper for purposes of comparison with the female vocal tract. In addition, only the vowels, /i/, /a/, and /u/ will be considered. The female subject (DJ) was a 27 year-old male with no history of speech or voice disorders and is native to the state of Texas in the southern United States. For comparison, the male subject (BS) was 29 years old at the time of scanning and is native to the midwestern United States. The male subject is also the first author of this study. Additionally, the vowels /i/ and /a/ were imaged for both subjects using EBCT. However, only the MRI-based images are discussed in this paper.

The MR images were acquired using a General Electric Signa 1.5 Tesla scanner. A 24 slice series of 5 mm thick contiguous, parallel, axial sections was gathered in an interleaved acquisition. This image set extended from just cephalad of the hard palate down to the first tracheal ring. The protocol for collecting the images consisted of having the subject positioned in a comfortable supine position on the patient table and was required to phonate at a comfortable pitch and loudness. The requirement of phonation was necessary in order for the subject to maintain the desired tract shape. More details concerning the scanning protocols can be found in Story[18] and Story et al.[17, 19, 20]

2.2. Image Analysis

All image analysis operations were performed with a general image display and quantitation package called VIDA (Volumetric Image Display and Analysis) which has been developed by researchers in the Division of Physiologic Imaging at the University of Iowa[21].

The image analysis process included three main steps: 1) segmentation of the airway from the surrounding tissue, 2) three-dimensional reconstruction of the airway by shape-based interpolation, and 3) determination of an airway centerline and subsequent extraction of cross-sectional areas assessed from oblique sections calculated to be locally perpendicular to the airway centerline to create an "area function". This image analysis process is identical to previous work which can be found in Story[18] and Story et al.[17, 19, 20] and also at the Internet web site - http://everest.radiology.uiowa.edu/tutor/app/vocal/vocal.html. Detailed explanations of the process will not be repeated here.

3. IMAGING RESULTS

The results of this experiment have been combined in Figures 1-3 with those of the male subject[17]. In Figure 1, the surface rendered representation of the 3-D vocal tract shape for the

Female Male

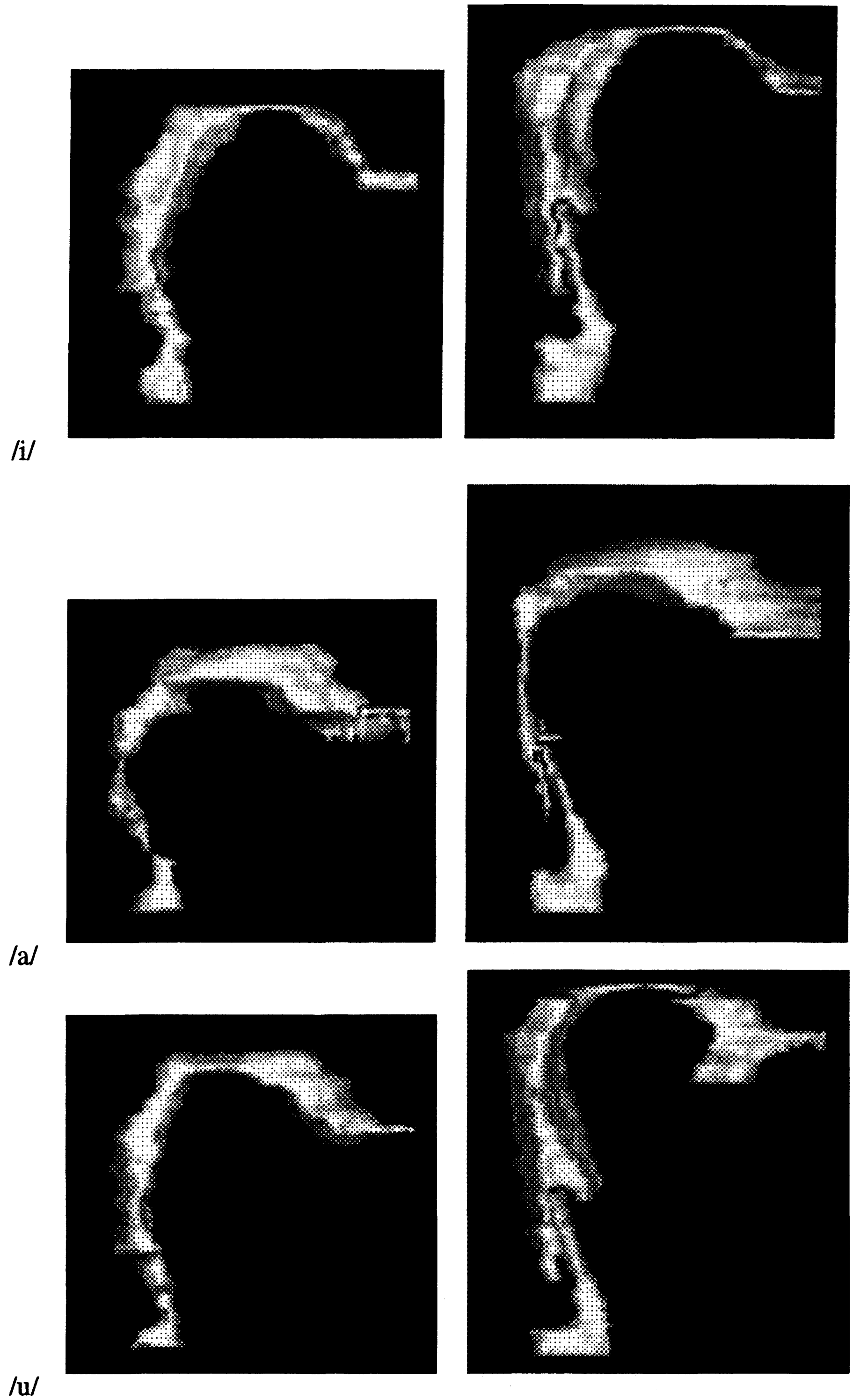

Figure 1: Shaded surface displays of three vowels (/i/, /a/, and /u/) for both a female (left column) and a male (right column).

female subject is shown on the left side of the figure and the male version on the right. For each surface rendered airway, the most inferior point of the 3-D shape begins with the uppermost section of the trachea. Above the trachea, the airway becomes small in the region of the glottis and then widens, more or less depending on the vowel, into the lower pharyngeal section. In the case of the male vocal tract shapes, the fingerlike extensions that hang down below the pharynx are the piriform sinuses. In the MR data sets for the female the piriform sinuses were poorly defined and consequently were not segmented. The area functions derived from the 3-D vowel shapes are given in Figure 2 for both male and female subjects. The 0 cm location represents a point just above the glottis and the termination at the lips is at right side of the graph. Figure 3 shows the vocal tract centerline profiles (a sagittal projection of the 3-D centerline) of each vowel for both male and female. In each graph, the point located at coordinate (0,0) represents a location just above the glottis; note that this point does not correspond to the lowest part of each surface rendered airway since those included the upper portion of the trachea. The profile then moves upward through the pharynx and then curves into the oral cavity and finally terminates at the lips. The solid dot located on each tract profile is the estimated location of the uvula tip; this is used as a landmark to divide the vocal tract into pharyngeal and oral sections.

The vocal tract can be further subdivided into an epilarynx which is defined, in this study, as the region of the vocal tract that extends from just above the glottis to the point where the piriform sinuses join the main vocal tract. This is an acoustic definition more than an anatomical one in that the widening of the vocal tract at the piriform junction creates an acoustic discontinuity that is important for lower vocal tract resonance and consequent voice quality characteristics[22]. It should be noted that Yang and Kasuya[5] used the tip of the epiglottis as an anatomical landmark to define this region. The solid dot on each area function in Figure 2 indicates the location of this point. Henceforth, the pharynx is defined as the section of the vocal tract between the end of the epilarynx and the tip of the uvula. The oral cavity extends from the uvula tip to the lips.

A general observation with regard to all three vowel shapes is that the overall size of the female vocal tract is, not surprisingly, smaller than that of the male. From both the surface rendered airways and the vocal tract profiles (Figs. 1 and 3) it is also observed that the female pharynx is typically shorter than that of the male but the oral cavities are of similar length. The solid dot indicating the uvula tip is assumed to divide the tract into pharyngeal and oral sections. The maximum cross-sectional areas shown in the area functions (Fig. 2) are generally smaller for the female than for the male.

The specific case of the area functions for vowel /i/ (Figure 2a) indicates that, for the female, the region just above the glottis linearly increases from an area of 0.2 cm^2 to 1.0 cm^2 at a point approximately 2 cm above the glottis. In comparison, the male cross-sectional areas in this same region remain nearly constant at about 0.3 cm^2. Beyond this point there is an abrupt increase in area for both male and female. The female tract reaches a maximum area of 4.5 cm^2 at a distance of 5 cm from the glottis after which the area begins to decrease until a minimum of 0.14 cm^2 is reached at 9.5 cm from the glottis. Following this minimum constriction, the area again rises to a value of 1.7 cm^2 at the lip termination which is 14.2 cm from the glottis. For the male vocal tract, a maximum area of 4.7 cm^2 occurs 7.1 cm from the glottis. As was the case for the female tract, the cross-sectional area following the maximum area begins to decrease and does so until an area of 0.1 cm^2 is reached at a point 11.8 cm from the glottis. The tract remains effectively constricted with areas in the range of 0.1 to 0.4 cm^2 until about 14 cm above the glottis at which point the area abruptly rises to 2 cm^2 close to the lip termination which is 16.2 cm from the glottis. The overall length of the vocal tract is 14.7 cm for the female and 16.1 cm for the male. Based on the definitions described above, the female epilarynx and pharynx length were measured to be 2.0 cm and 4.2 cm, respectively while the oral cavity was 8.5 cm

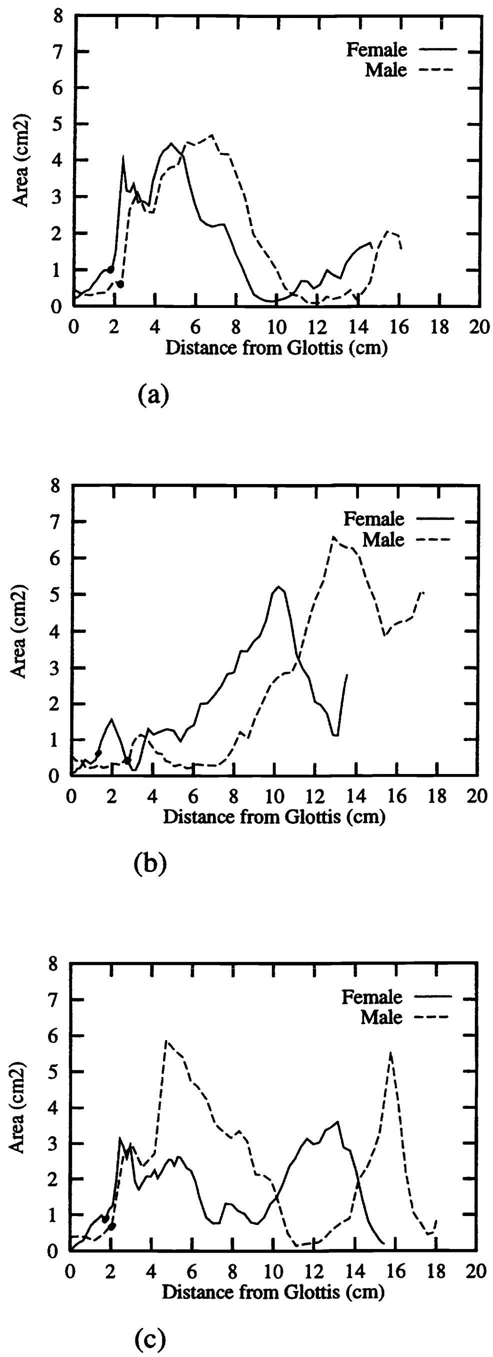

Figure 2: Area functions for female (solid) and male (dashed) vowels. The solid dot on each area function indicates the point at which the piriform sinuses join the main vocal tract. a) /i/, b) /a/, and c) /u/.

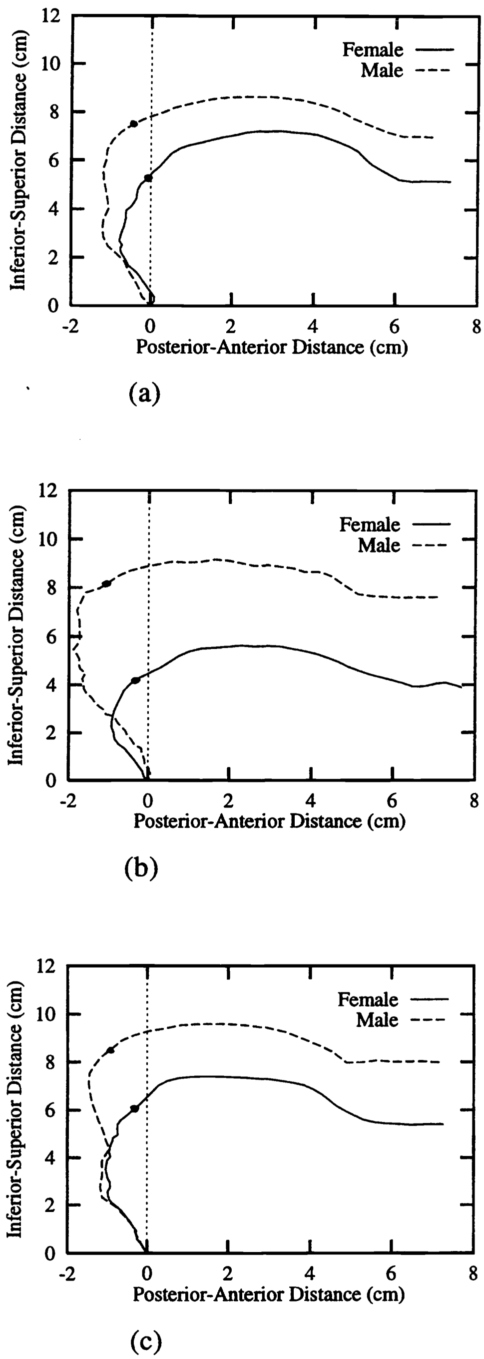

Figure 3: Sagittal projections of the vocal tract centerline for female (solid) and male (dashed) vowels. The solid dot on each projection indicates the location of the uvula tip. a) /i/, b) /a/, and c) /u/.

long. For the male, the epilarynx, pharynx and oral cavity lengths were measured to be 2.4 cm, 6.0 cm, and 7.7 cm, respectively.

The area function for the female vowel /a/ (Figure 2b) shows that the area gradually increases from 0.05 cm^2 just above the glottis to about 0.4 cm^2 at 1 cm from the glottis. The area then abruptly rises to a peak of 1.7 cm^2 at 2 cm above the glottis and then, just as abruptly as it rose, the area falls to 0.15 cm^2 at a point 3 cm above the glottis. Following this constriction, the area sharply increases to 1.3 cm^2 and remains nearly constant for a 1.5 cm long section of the tract. At a point about 5.3 cm above the glottis, the area begins to increase into the widened oral cavity characteristic of the /a/ vowel. A maximum area of 5.1 cm^2 occurs at 10 cm above the glottis after which the area decreases to 1.5 cm^2 before rising to 2.8 cm^2 at the lip termination. The lower pharyngeal section (or the epilarynx) of the male /a/ is quite different from the female in that a nearly constant area of approximately 0.25 cm^2 is observed from just above the glottis up to a location 2.5 cm above the glottis. Then a sharp rise in area occurs with a peak of 1.1 cm^2 located 3.4 cm above the glottis. The area gradually drops to 0.25 cm^2 before increasing from about the 7 cm point up to the 13.5 cm point reaching a peak value of 6.6 cm^2. From the point of peak area out to the mouth termination the area function decreases before rising to lip termination area of 5.0 cm^2. The vocal tract length is 13.6 cm for the female and 17.3 cm for the male. Based on the locations of the uvula tip and piriform sinus junction, the female epilarynx, pharynx and oral cavity lengths were found to be 1.5 cm and 3.0 cm while measurements for the male gave an epilarynx length of 2.8 cm, a pharynx length of 5.9 cm and oral cavity length of 8.6 cm.

The vowel /u/ for the female has an area function that begins much like the /a/ vowel with an small area of 0.05 cm^2 just above the glottis. The area then increases gradually to a value of 1.1 cm^2 before increasing abruptly to 3.1 cm^2 at a distance of 2.2 cm from the glottis. Over the course of the next 4.5 cm the area falls into the range of 2.5 cm^2 but retaining the characteristic pharyngeal expansion of an /u/. The mid-tract constriction occurs from 7 cm to 9 cm with a minimum area in this region of 0.9 cm^2. The area then rises to a peak of 3.5 cm^2 before falling to 0.2 cm^2 at the lip termination. Much like the previous two vowels, the epilaryngeal region for the male /u/ has a nearly constant area, in this case about 0.38 cm^2. This is in contrast to the female vowel where the epilarynx has a more horn-like shape. Following the epilarynx, the male /u/ has an rapid increase in area which climbs to a maximum of 6.0 cm^2 the 4.7 cm point. The area gradually decreases over the next 6.0 cm until a mid-tract constriction area of 0.15 is reached at the 11 cm point. This is considerably smaller constriction area than for the female where 0.9 cm^2 was the minimum. The constriction is approximately 2 cm long after which the area increases into the front cavity expansion where a maximum area of 5.5 cm^2 occurs. The area then drops to 0.7 cm^2 at the lip termination. The overall length of the vocal tract is 15.5 cm for the female and 18.0 cm for the male. The length of the female epilarynx was measured to be 1.9 cm, the pharynx 2.0 cm, and the oral cavity 8.7 cm. The male epilarynx was 2.0 cm long while the pharynx and oral cavity had a length of 7.1 cm and 8.9 cm, respectively.

The measurements of the total vocal tract length, lengths of the epilaryngeal, pharyngeal and oral sections and maximum cross-sectional areas are summarized in Table 1. The pharynx lengths for the female range from 1.8 cm to 3.0 cm shorter than the male while the oral cavity lengths are similar across gender. For the epilarynx lengths, vowels /i/ and /u/ are similar for the both male and female whereas the female epilarynx in the vowel /a/ is nearly half that of the corresponding male vowel. The maximum areas are, in general, slightly larger for the male than the female.

Table 1: Lengths of the epilarynx, pharynx, and oral cavities for the female and male vowels /i/, /a/, and /u/. The maximum area measured in each area function is also given. All measurements are in centimeters unless otherwise indicated.

	Female			*Male*		
Cavity	**/i/**	**/a/**	**/u/**	**/i/**	**/a/**	**/u/**
epilarynx	2.0	1.5	1.9	2.4	2.8	2.0
pharynx	4.2	3.0	4.8	6.0	5.9	7.1
oral	8.5	9.2	8.7	7.7	8.6	8.9
max area(cm^2)	4.45	5.23	3.60	4.69	6.57	5.87

In Table 2, ratios of the female measurements to those of the male are presented for the three vowels and a mean is computed across vowels. The ratios show that all regions of the female vocal tract are shorter than the male except the oral cavity. In fact, the female oral cavity is slightly larger than the male for /i/ and /u/ and nearly the same for the /a/. This suggests that a female vocal tract is not simply a compressed version of the male vocal tract but is nonuniformly adjusted across various regions with respect to a male. The set of ratios for a given vowel provides a kind of "recipe" for transforming a male vocal tract shape into a female vocal tract or vice versa; the set of mean ratios gives a generalized "recipe". In the average sense, an vocal tract area function for the male subject could be approximately transformed into a female-like area function by compressing the epilarynx to 77 percent of its original value, the pharynx to 63 percent of its original, and expanding the oral cavity by 5 percent. The words "compress" and "expand" are deliberately used instead of "shorten" and "lengthen" to stress that the shape of the area function must be retained throughout the vocal tract; e.g. the various regions are not simply truncated to realize a shorter section length. If only the total vocal tract length were considered as the main contributor to male versus female quality, the female vocal tract would be realized by compressing the entire male vocal tract by 15 percent, on the average. The ratios in this table are quite similar to those given in Goldstein[4] , Yang and Kasuya[5] and Hogberg[6].

Table 2: Female to male ratios of the vocal tract region lengths and maximum areas given in Table 2.

Female/Male	**/i/**	**/a/**	**/u/**	**mean**
epilarynx	0.83	0.54	0.94	0.77
pharynx	0.70	0.51	0.68	0.63
oral	1.10	1.07	0.98	1.05
total length	0.91	0.79	0.86	0.85
max area	0.95	0.79	0.61	0.78

4. ACOUSTIC MODELING OF MALE AND FEMALE VOCAL TRACT SHAPES

Acoustic resonance characteristics of each male and female area function were computed with a frequency domain transmission line[23]. The results are shown in Figure 4. In each plot, the thick line indicates the formant spectrum for the female vowel shape while the thinner line represents the male; each peak in the spectrum is called a formant and it is the pattern of formants that defines a specific vowel or consonant.

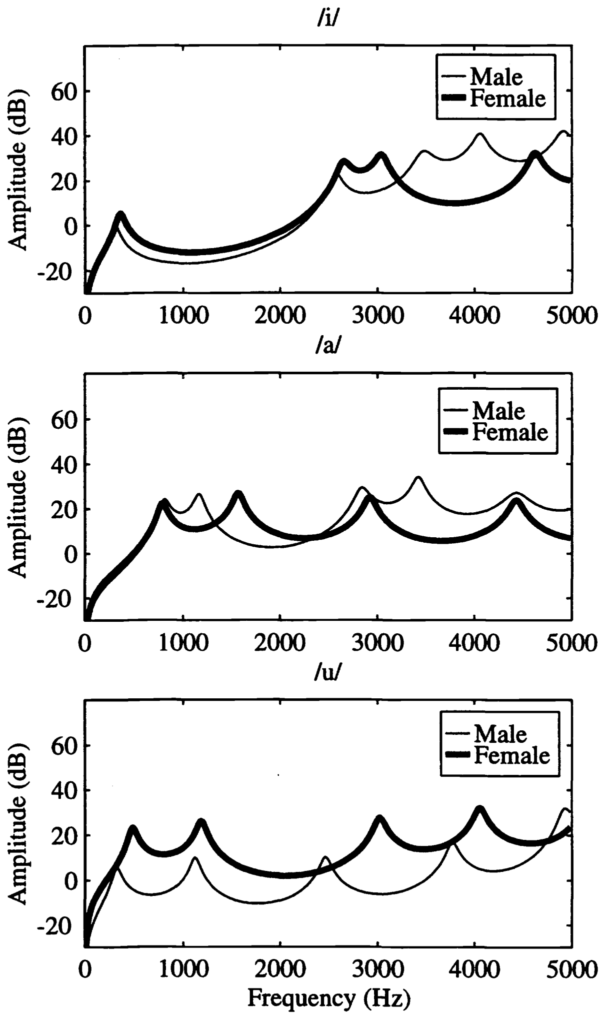

Figure 4: Formant spectra of vowels /i/, /a/, and /u/ for female (thick lines) and male (thin lines) subjects.

For the vowel /i/, the female F1 is slightly higher that the F1 of the male. F2 and F3 of the female are clustered together in the 2800 Hz to 3100 Hz range such that F2 for the female is higher than the male and F3 is lower than for the male. The female F4 is located at about 4700 Hz which gives an F4-F3 difference of approximately 1600 Hz, while the difference of F4-F3 for the male is only 700 Hz. In fact, the frequency range of 3000 Hz to 4500 Hz is highly accentuated for the male but is of relatively low amplitude for the female. This observation is similar to that described by Agren and Sundberg[9] for alto and tenor voices. For the /a/, the female F1 is slightly lower than the male but F2 is about 300 Hz higher. F3 is similar for both male and female but the female F4 appears at 4200 Hz, about 900 Hz higher than F4 for the male. The formant spectra for the /u/ shows higher formant locations for the female than the male across the entire frequency spectrum.

An attempt was made to nonuniformly transform the male /i/ vowel into a female /i/ by applying the ratios calculated in Table 2. Thus, the epilarynx length, was compressed by 17 percent, the pharynx length compressed by 30 percent, and the length of the oral cavity expanded by 10 percent. The original female area function for /i/ is shown in Figure 5a as the thin solid line while the transformed male-to-female area function is shown with a thick solid line. Note that the areas in the male /i/ have not been changed, only the lengths of the three vocal tract regions have been altered. The original female and transformed male area functions are quite similar in that the points of expansion and constriction occur in nearly the same locations. The cross-sectional area throughout the length of the entire area function are also very similar. The formant spectra of each of the area functions in Figure 5a is given in Figure 5b with the same line styles. The transformed male area function is observed to have formant characteristics much like the female; F1 is nearly the same as the female, F2 and F3 are similarly clustered although situated about 400 Hz higher in frequency, and the F4's are located at nearly the same frequency. Thus, the male vocal tract appears to have been transformed into a vocal tract with female characteristics.

Another attempt to transform the male area function for /i/ into one with female-like qualities was performed by uniformly compressing the length of the entire area function by 9 percent. (see Table 2). Shown in Figure 6a is the original female area function for /i/ and the male area function uniformly scaled to female length. Unlike, the previous nonuniform transformation, the expansions and constrictions of the transformed male to female area function do not match closely with the original female /i/ area function. The transformed area function is shifted to the right relative to the original female area function. The formant spectra for the original female /i/, the previous (nonuniform) transformation, and this simple length scaling are shown in Figure 6. The formant spectrum from the uniform length scaling appears to have the same general shape as the original male /i/ vowel spectrum, except shifted upward in frequency by a small amount. This would imply that possibly the male characteristics of the vocal tract have been retained even though the formant locations have been moved up in frequency. However, it should also be noted that the locations of F1 and F2 are closer to those of the female /i/ spectrum than are those from the nonuniform scaling of the male /i/. Thus, if the absolute locations of F1 and F2 are important for female voice quality then the uniform scaling may the appropriate choice. Conversely, if the overall pattern of formant clustering characteristics is crucial to realizing a female quality, then the nonuniform scaling needs to be used. This would be especially true if Agren and Sundberg's[9] hypothesized significance of the distance between F3 and F4 is correct. A formal perceptual study would need to be performed to determine whether uniform or nonuniform scaling should be used to evoke a female-like quality. However, an informal listening test has suggested that the nonuniformly scaled /i/ vowel produces a slightly more female-like quality than the uniformly scaled version.

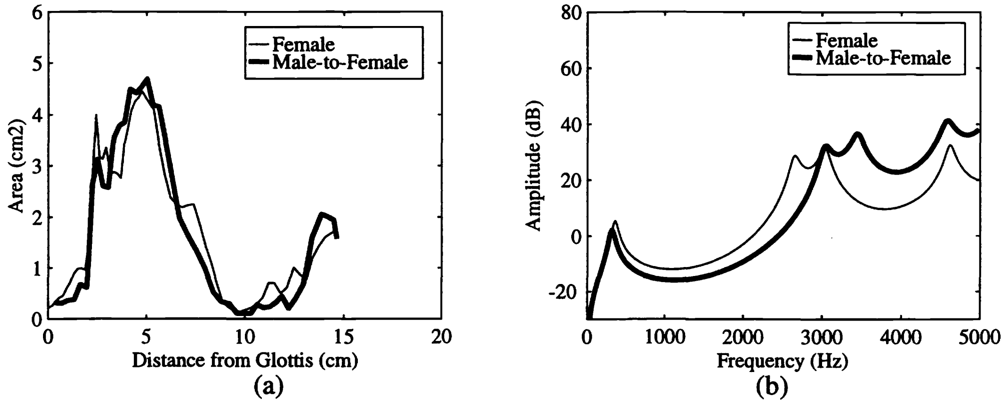

Figure 5: Original female /i/ and nonuniform transformation from male /i/ to female /i/; a) area functions, b) vocal tract transfer functions.

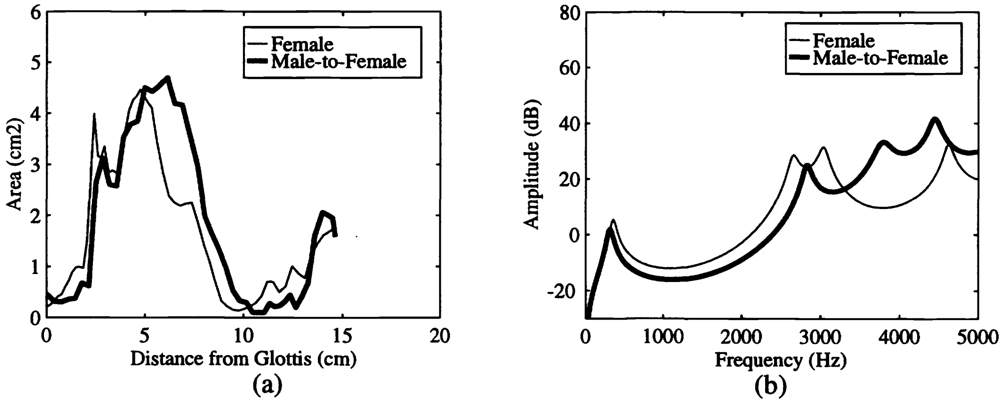

Figure 6: Original female /i/ and uniform transformation of male /i/ to female /i/; a) area functions, b) vocal tract transfer functions.

5. CONCLUSIONS

3-D vocal tract shapes representing three vowels (/i/, /a/, and /u/) have been acquired from one male and one female subject using MRI. The 3-D vocal tract shapes were condensed into an area function which was divided into three distinct regions representing the epilarynx, pharynx, and oral cavity. The epilarynx and pharynx were shorter than those of the male subject while the female oral cavity was nearly the same length or even slightly longer than for the male. Ratios of the female to male lengths of these three regions produced a simple recipe for transforming a male vocal tract shape to one with female qualities or vice versa. It was observed that, for the /i/ vowel, the nonuniform transformation was necessary in order to obtain the spectral qualities representative of a female. A simple uniform compression of the area functions shifted the formants higher in frequency, it did not reorganize the relationship of the formants to one another. This suggests that a female voice quality is largely dependent upon a shortened pharynx. This study has only begun to explore the male versus female vocal tract differences for vowels and consonants but based on the limited information presented it appears that a female vocal tract shape contains unique qualities that cannot be explained by a simple uniform compression of the male vocal tract.

6. ACKNOWLEDGMENTS

This study was supported by grant No. R01 DC02532-02 from the National Institutes on Deafness and Other Communication Disorders.

7. REFERENCES

1. Wu, K., and Childers, D.G., "Gender recognition from speech. Part I: Coarse analysis," J. Acoust. Soc. Am., 90(4), 1828-1840, 1991.

2. Klatt, D.H., and Klatt, L.C., "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am., 87(2), 820-857, 1990.

3. Titze, I.R., "Physiologic and acoustic differences between male and female voices," J. Acoust. Soc. Am., 85(4), 1699-1707, 1989.

4. Goldstein, U.G., "An articulatory model for the vocal tracts of growing children," Ph.D. Dissertation, MIT, 1980.

5. Yang, C.S., and Kasuya, H., "Uniform and nonuniform normalization of vocal tracts measured by MRI across male, female, and child subjects," IEICE Trans. Inf. and Sys., E78-D (6), 732-737, 1995.

6. Hogberg, J., "From sagittal distance to area function and male to female scaling of the vocal tract," STL-QPSR, 4, 11-66, 1995.

7. Wong, D., Lange, R.C., Long, R.K., Story, B.H., and Titze, I.R., "Age and gender related speech transformations using linear predictive coding," J. Acoust. Soc. Am., (in review), 1997.

8. Coleman, R.O., "A comparison of the contributions of two vocal characteristics to the perception of maleness and femaleness in the voice," STL-QPSR, 2-3, 13-22, 1973.

9. Agren, K., and Sundberg, J., "An acoustic comparison of alto and tenor voices," STL-QPSR, 1, 12-16, 1976.

10. Terhardt, E., "On the perception of periodic sound fluctuations (roughness)", Acustica, 30, 201-213, 1974.

11. Ingemann, F., "Identification of the speaker's sex from voiceless fricatives," J. Acoust. Soc. Am., 44, 1142, 1968.

12. Schwartz, M., "Identification of speaker sex from isolated voiceless fricatives," J. Acoust. Soc. Am., 43, 1178, 1968.

13. Baer, T., Gore, J. C., Gracco, L. C., and Nye, P. W., "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," J. Acoust. Soc. Am., 90, 799-828, 1991.

14. Dang, J., Honda, K., and Suzuki, H., "Morphological and acoustical analysis of the nasal and the paranasal cavities," J. Acoust. Soc. Am., 96(4), 2088-2100, 1994.

15. Narayanan, S. S., Alwan, A. A., and Haker, K., "An articulatory study of fricative consonants using magnetic resonance imaging," J. Acoust. Soc. Am., 98(3), 1325-1347, 1995.

16. Narayanan, S. S., "Fricative consonants: An articulatory, acoustic, and systems study," Ph. D. thesis, UCLA, Dept. of Electrical Engineering, Los Angeles, CA.

17. Story, B.H., Titze, I.R., and Hoffman, E.A., "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am., 100(1), 537-554, 1996.

18. Story, B. H., "Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract," Ph. D. Dissertation, University of Iowa, Dept. of Speech Pathology and Audiology, 1995.

19. Story, B. H., Hoffman, E. A., and Titze, I. R., "Speech simulation based on MR images of the vocal tract," SPIE Proc. Physiology and Function from Multidimensional Images, 2433, San Diego, CA, 26 Feb. - 2 Mar., 1995.

20. Story, B. H., Hoffman, E. A., and Titze, I. R., "Vocal tract imaging: A comparison of MRI and EBCT," SPIE Proc. Physiology and Function from Multidimensional Images, 2709, 209-222, Newport Beach, CA, 11-13 February, 1996.

21. Hoffman, E. A., Gnanaprakasam, D., Gupta, K. B., Hoford, J. D., Kugelmass, S. D., and Kulawiec, R. S., "VIDA: An enviroment for multidimensional image display and analysis," SPIE Proc. Biomed. Image Proc. and 3-D Microscopy, 1660, San Jose, CA, 10-13 Feb., 1992.

22. Titze, I.R., and Story, B.H., "Acoustic interactions of the voice source with the lower vocal tract," J. Acoust. Soc. Am., (in press), 1997.

23. Sondhi, M.M., and Schroeter, J., "A hybrid time-frequency domain articulatory speech synthesizer," IEEE Trans. ASSP, ASSP-35(7), 1987.