127 W

Posterior Pharyngeal Wall Position in the Production of Speech

Harriet S. Magen
Haskins Laboratories
New Haven, CT
and
Rhode Island College
Providence

A. Min Kang
Haskins Laboratories
and
Yale University
New Haven, CT

Mark K. Tiede
Haskins Laboratories
New Haven, CT
and
Massachusetts Institute
of Technology
Cambridge

D. H. Whalen
Haskins Laboratories
New Haven, CT

The posterior pharyngeal wall has been assumed to be stationary during speech. The present study examines this assumption in order to assess whether midsagittal widths in the pharyngeal region can be inferred from measurements of the anterior pharyngeal wall. Midsagittal magnetic resonance images and X-ray images were examined to determine whether the posterior pharyngeal wall from the upper oropharynx to the upper laryngopharynx shows anterior movement that can be attributed to variables in speech: vowel quality in both English and Japanese; vowels versus consonants as classes of speech sounds; sustained versus dynamically produced speech; and isolated words versus sentences. Measurements were made of the distance between the anterior portion of the vertebral body and the pharyngeal wall. The first measurement was on a line traversing the junction between the dens and the body of the second cervical vertebra (C2). The next three measurements were on lines at the inferior borders of the bodies of C2, C3, and C4. The measurements showed very little movement of the posterior pharyngeal wall, none of it attributable to speech variables. Therefore, the position of the posterior pharyngeal wall in this region can be eliminated as a variable, and the anterior portion of the pharynx alone can be used to estimate vocal cavities.

KEY WORDS: pharyngeal wall, speech production, vocal tract, magnetic resonance imaging, X-ray cineradiography

The pharynx is a muscular tube extending from the base of the skull to the larynx. Three pairs of constrictor muscles (superior, middle, and inferior) circle it along its length. Having no muscular attachments to the vertebrae, it is capable of independent movement. Nevertheless, the posterior wall below the nasopharynx has been assumed to be virtually motionless during speech (Carmody, 1941; Iglesias, Kuehn, & Morris, 1980; Westbury, 1983; Zemlin, 1998). This assumption is based on data derived from conventional X rays (Moll, 1960) and therefore limited with respect to both interpretation and data collection because of problems associated with imaging soft tissues as well as those associated with exposure to radiation (see Whalen, Kang, Magen, Fulbright, & Gore, 1999, for further discussion). Moreover, because of its importance in the attainment of velopharyngeal closure, the region of the pharynx measured in previous studies of speech production has been mostly limited to the oropharynx and above, whereas study of the area below the oropharynx has been lacking.

The posterior pharyngeal wall in speech has been of interest to researchers largely for the role it exerts in the attainment of velopharyngeal closure, whereby the velopharyngeal mechanism varies the degree of acoustic coupling between oral and nasal cavities. In normal individuals, velopharyngeal closure is achieved by elevating and retracting the soft palate and at the same time constricting the walls of the nasopharynx without movement of the posterior pharyngeal wall. In some individuals, however, particularly those with a short palate, such as those with cleft palate, there appears to be compensatory movement of the upper posterior pharyngeal wall anteriorly to meet the soft palate (e.g., Zemlin, 1998). To assess the difference in velopharyngeal closure between normal and postoperative cleft palate patients, Hagerty and colleagues performed extensive laminographic X-ray studies of the velopharyngeal region in which participants were studied under three conditions: rest, phonating [ɑ], and producing [s] (Hagerty & Hill, 1960; Hagerty, Hill, Pettit, & Kane, 1958). Results indicated that although there is greater pharyngeal wall movement for [s] than for [ɑ] in normal participants, the actual forward movement of the posterior pharyngeal wall is not necessary for speech (Hagerty et al., 1958). For postoperative cleft palate patients, on the other hand, there is a tendency for more movement of the pharyngeal wall to compensate for reduced velar movement in the attainment of velopharyngeal contact (Hagerty & Hill, 1960). This is accompanied by a greater incidence among the cleft palate group of a Passavant's pad, an acquired transverse swelling or bulge of muscle on the posterior wall of the pharynx at the level of the base of the uvula, which facilitates velopharyngeal closure (Hagerty & Hill, 1960; Harrington, 1944; Passavant, 1863; Seikel, King, & Drumright, 2000) (cf. Calnan, 1954).

The pharynx wall in the oropharyngeal region has also been studied for its role in the production of consonant voicing distinctions. Based on cineradiographic data, Perkell (1969) and Kent and Moll (1969) reported sagittal oropharyngeal width to be consistently larger for voiced than for voiceless consonants, resulting in a pressure drop that allows glottal pulsing to continue during vocal tract occlusion. However, neither of these studies differentiated the contribution of the tongue as distinct from the posterior pharyngeal wall in their measures of oropharyngeal width. EMG results of superior and middle pharyngeal constrictor muscles indicated less EMG activity during voiced consonants, thereby allowing the increase in volume (Bell-Berti, 1975; Minifie, Abbs, Tarlow, & Kwaterski, 1974). These data showed greater EMG activity for the vowels examined, [i, ɑ], than for the consonants examined, [b, p], but no difference in degree of activity between the two vowels (Minifie et al., 1974).

In contrast to EMG data indicating no difference in degree of muscular activity depending on vowel quality, such differences have been found both in movement of the lateral pharyngeal wall using ultrasound data (Minifie, Hixon, Kelsey, & Woodhouse, 1970; Zagzebski, 1975) and velopharyngeal closure in vowels using a cinefluorographic technique (Moll, 1962). The extent of inward movement of the lateral pharyngeal wall has been shown to vary according to vowel height, with the greatest movement during low vowels and the least during high vowels (Minifie et al., 1970); low vowels exhibited less velopharyngeal closure than high vowels (Moll, 1962). Although lateral pharyngeal wall movement at the level of the middle pharyngeal constrictors exhibited the high versus low difference, lateral pharyngeal wall movement at the level of the superior constrictors exhibited a nasal versus non-nasal difference, with greater motion for non-nasal vowels (Zagzebski, 1975).

In addition to its role in speech, the posterior pharyngeal wall participates in the execution of involuntary actions, principally swallowing (deglutition) and gagging; movements of the posterior pharyngeal wall, including those below the oropharynx, have therefore been studied in this context. These data indicate that swallowing disorders frequently occur in patients whose swallows are characterized by relative immobility in this region (Cunningham, Donner, Jones, & Point, 1991). A videoradiographic study of swallowing, using radiopaque markers affixed to the pharyngeal wall by suction, showed marker movement of 4 to 7 mm in the anterior-posterior direction in the oropharyngeal area (Palmer, Tanaka, & Siebens, 1988). Similarly, a videofluoroscopic study of an oropharyngeal measurement point during swallowing, studying healthy men in two age groups, showed anterior movement of the posterior pharyngeal wall on the order of 6–7 mm in both groups (Logemann et al., 2000). In an EMG study comparing participants on several reflexive and nonreflexive tasks, reflexive tasks, such as swallowing and gagging, produced significantly higher EMG amplitudes of the superior pharyngeal constrictor than did the speech tasks studied (Perlman, Luschei, & DuMond, 1989). Posterior pharyngeal wall movements of this magnitude during speech would be expected to have acoustic consequences.

In this study, we examine posterior pharyngeal wall position from upper oropharynx to upper laryngopharynx, supplementing the study of the oropharynx and above, the area of the pharynx that has received extensive study. We use two imaging techniques, magnetic resonance imaging (MRI) and X-ray cineradiography, each of which offers advantages while suffering certain disadvantages in process and result. The advantages of MRI over X ray are that it spares the participant exposure to radiation and that shadows resulting from projecting the three-dimensional vocal tract onto a two-dimensional film are avoided. For our purposes, the two disadvantages of MRI are that acquisition times are slow, necessitating sustained productions, and that the participant must be in a supine position, where the different direction of gravitational loading possibly alters the production of vowels (Tiede, Masaki, & Vatikiotis-Bateson, 2000). Although the effects of gravity are apparent in articulations involving the jaw, a relatively massive articulator (Shiller, Ostry, & Gribble, 1999; Tiede et al., 2000), gravity is less likely to affect the pharyngeal wall because of its thinness. The advantages of X-ray imaging are that the participant can sit in normal phonation posture and that dynamic images of the vocal apparatus can be collected. These advantages are offset by the disadvantages of exposing the participant to radiation during the process and the shadows on the images. It is hoped that the combination of magnetic resonance (MR) and X-ray images will result in a more complete picture of the pharyngeal region of interest.

The aim of the present study is to determine the amount and location of posterior pharyngeal wall movement in speech that is not attributable to velopharyngeal closure and to determine the extent to which this movement is predictable. One goal at Haskins Laboratories is vocal tract modeling, which we believe can be accomplished by ultrasonic measurement of the anterior portion of the pharynx. If the position of the posterior pharyngeal wall from upper oropharynx to upper laryngopharynx is predictable and can therefore be eliminated as a variable, the anterior portion of the pharynx alone could be used to estimate vocal cavities. In this study, we investigate possible differences in posterior wall position, not only among oral vowels of various qualities, but also between vowels and consonants as classes of speech sounds. Our interest is in the vocal tract region extending from the upper oropharynx to upper laryngopharynx; therefore, we do not examine the nasal/non-nasal contrast. We assess possible differences between the two methods of production—sustained productions and running speech—that are necessitated by the MR and X-ray imaging techniques, respectively. Within running speech contexts, we explore possible differences between isolated words and sentences. Our data include productions from native speakers of English as well as native speakers of Japanese.

Experiment 1: MRI

Method

MR images were collected at two sites: Yale University School of Medicine in New Haven, Connecticut, and the Takanohara Central Hospital in Nara, Japan.

Participants

A total of 13 adults participated: 7 native speakers of American English (5 male, 2 female) and 6 native speakers of Japanese (5 male, 1 female). Data from 2 of the American English speakers (1 male, 1 female) were collected in the United States; all other data were collected in Japan. The English speakers ranged in age from 29 to 46 and the Japanese speakers from 30 to 54. The 2 English speakers scanned in New Haven and 2 of the English speakers scanned in Japan spoke dialects that preserve the /ɑ/-/ɔ/ distinction. All speakers were neurologically normal with no history of speech or hearing disorders.

Stimuli

Nine oral vowels were collected from all American English speakers: [i ɪ eʲ ɛ æ ɑ oʷ ʊ u], and [ɔ] was collected from those speakers who make the /ɑ/-/ɔ/ distinction. In addition, [ʌ] was collected from the two speakers scanned in the United States. Complete inventories of the Japanese five-vowel system, [i e a o ɯ], were collected from the Japanese speakers. Acquisition time varied according to the imaging protocol used (see the Apparatus section below). Participants were instructed to vocalize and hold the initial position of each vowel for the duration of acquisition, taking shallow breaths when necessary and inverting airflow through the tract without changing articulator position. Participants' vocalizations were monitored over an intercom system for any obvious deviations induced by swallowing or respirations; images showing motion artifacts were repeated. The number and type of tokens collected from each English-speaking and Japanese-speaking participant are indicated in parentheses in Tables 1 and 2, respectively.

Apparatus

In the United States, MR images were collected on a General Electric 1.5T Signa magnet. Sagittal images (5-mm thickness) were acquired using a fast spin-echo technique with the following parameters: repetition time (TR) = 400 ms, echo time (TE) = 14 ms, echo train length (ETL) = 4, echo spacing (Esp) = 14 ms, 128 × 128, number of excitations (nex) = 1, field of view (FOV) = 28 cm. Acquisition times were about 14 s. Complete scanning procedures for the New Haven data have been reported elsewhere (Whalen et al., 1999).

In Japan, MR images were obtained on a Shimadzu 1T magnet (SMT-100GUX). For 3 of the Japanese-speaking participants (JF1, JM1, JM2), contiguous sagittal images (3-mm thickness) were collected using a conventional spin-echo technique with the following parameters: TR = 300 ms, TE = 18 ms, 256 × 256, nex = 2,

Table 1. Mean distances (mm ± *SD*; no *SD* for single tokens) at each measurement level for the 7 English speakers.

Speaker	Level	[i]	[ɪ]	[eʲ]	[ɛ]	[æ]	[ɑ]	[ɔ]	[ʌ]	[oʷ]	[ʊ]	[u]
EF1[a]	2	3.6 ± 0.9 (6)	3.9 ± 0.9 (6)	4.2 ± 0.4 (6)	3.9 ± 0.8 (4)	4.4 ± 0.4 (6)	3.9 ± 0.5 (6)	3.8 ± 0.3 (6)	3.9 ± 0.4 (6)	3.7 ± 0.9 (6)	4.3 ± 0.5 (6)	3.7 ± 0.6 (6)
	3	3.6 ± 0.8 (6)	4.6 ± 0.5 (6)	4.6 ± 1.1 (6)	4.5 ± 0.4 (4)	4.8 ± 0.4 (6)	4.6 ± 0.8 (6)	4.9 ± 0.9 (6)	4.3 ± 0.3 (6)	4.5 ± 0.4 (6)	4.6 ± 0.7 (6)	4.1 ± 0.6 (6)
EF2[b]	2	2.6	2.5	1.8	2.0		2.3			1.8	2.3	1.9 ± 0.1 (2)
	3	2.4	2.6	2.3	2.4		2.9			2.6	2.1	2.2 ± 0.2 (2)
EM1[a]	1	1.4 ± 0.5 (5)	2.9 ± 0.3 (5)	2.9 ± 0.7 (5)	1.9 ± 0.5 (5)	2.3 ± 0.9 (5)	3.1 ± 0.8 (5)	2.7 ± 0.9 (5)	1.5 ± 0.9 (5)	2.0 ± 0.8 (5)	1.3 ± 0.7 (5)	2.0 ± 1.2 (5)
	2	1.8 ± 0.5 (5)	3.1 ± 0.4 (5)	2.1 ± 0.9 (5)	2.8 ± 0.4 (5)	1.9 ± 0.8 (5)	2.3 ± 0.7 (5)	2.3 ± 0.7 (5)	2.6 ± 0.6 (5)	2.3 ± 0.4 (5)	2.1 ± 0.8 (5)	1.9 ± 0.6 (5)
	3	3.0 ± 0.8 (5)	4.1 ± 0.5 (5)	2.7 ± 0.8 (5)	3.6 ± 0.3 (5)	3.3 ± 0.3 (5)	2.8 ± 0.9 (5)	3.0 ± 0.3 (5)	2.8 ± 0.3 (5)	2.7 ± 0.4 (5)	3.3 ± 0.4 (5)	2.6 ± 0.4 (5)
	4	2.6 ± 1.3 (5)	3.4 ± 0.5 (5)	3.0 ± 0.7 (5)	3.2 ± 0.5 (5)	2.5 ± 0.9 (5)	3.5 ± 0.9 (5)	2.9 ± 0.4 (5)	2.1 ± 0.2 (5)	3.3 ± 0.4 (5)	3.0 ± 0.3 (5)	3.0 ± 0.5 (5)
EM2[b]	2	2.1	2.6		2.1	2.3	2.1 ± 0.4 (2)			2.2	2.0	1.8
	3	1.8	2.3		2.0	1.8	2.0 ± 0.4 (2)			1.8	2.0	1.8
	4	2.5	2.8		2.8	2.0	2.9 ± 0.2 (2)			2.5	2.3	2.3
EM3[b]	2	2.4 ± 0.2 (2)	2.5 ± 0.4 (2)	3.2 ± 0.2 (2)	3.3 ± 0.0 (2)	2.7 ± 0.2 (2)	2.8 ± 0.1 (2)			2.5	2.5	2.3
	3	2.4 ± 0.5 (2)	2.9 ± 0.2 (2)	2.9 ± 0.2 (2)	2.8 ± 0.4 (2)	1.9 ± 0.2 (2)	2.3 ± 0.4 (2)			2.1	2.6	2.3
	4	3.0 ± 0.4 (2)	3.5	3.0 ± 0.0 (2)	3.8	2.8 ± 0.4 (2)	3.3			3.5	2.5	
EM4[b]	2	2.0 ± 0.2 (4)	2.6 ± 0.2 (3)	2.1 ± 0.2 (3)	2.4 ± 0.1 (3)	2.1 ± 0.2 (3)	2.1 ± 0.0 (2)	2.2 ± 0.5 (2)		2.3 ± 0.2 (3)	2.1	2.4 ± 0.2 (5)
	3	2.5 ± 0.3 (4)	3.1 ± 0.0 (3)	2.6 ± 0.2 (3)	2.8 ± 0.2 (3)	2.5 ± 0.5 (3)	2.9 ± 0.2 (2)	2.5 ± 0.0 (2)		2.4 ± 0.0 (3)	2.8	2.3 ± 0.3 (5)
	4	3.3 ± 0.3 (4)	4.1 ± 0.3 (3)	3.8 ± 0.4 (3)	3.9 ± 0.4 (3)	3.8 ± 0.5 (3)	3.3 ± 0.4 (2)	3.0 ± 0.4 (2)		3.4 ± 0.2 (3)	3.3	2.5 ± 0.1 (5)
EM5[b]	2	3.3 ± 0.1 (2)	2.6 ± 0.4 (2)	2.9	3.1	2.7 ± 0.2 (2)	2.7 ± 0.2 (2)	2.2 ± 0.1 (2)		2.4 ± 0.6 (2)	2.8	2.5
	3	3.5 ± 0.0 (2)	2.7 ± 0.1 (2)	3.3	3.4	2.8 ± 0.0 (2)	2.7 ± 0.2 (2)	2.7 ± 0.2 (2)		2.3 ± 0.4 (2)	2.8	2.8
Mean	1	1.4	2.9	2.9	1.9	2.3	3.1	2.7	1.5	2.0	1.3	2.0
	2	2.5 ± 0.7	2.8 ± 0.5	2.7 ± 0.9	2.8 ± 0.7	2.7 ± 0.9	2.6 ± 0.6	2.6 ± 0.8	3.2 ± 0.9	2.4 ± 0.6	2.6 ± 0.8	2.4 ± 0.7
	3	2.7 ± 0.7	3.2 ± 0.8	3.1 ± 0.8	3.1 ± 0.8	2.8 ± 1.1	2.9 ± 0.8	3.3 ± 1.1	3.5 ± 1.1	2.6 ± 0.9	2.9 ± 0.9	2.6 ± 0.7
	4	2.9 ± 0.3	3.5 ± 0.5	3.3 ± 0.4	3.4 ± 0.5	2.8 ± 0.9	3.1 ± 0.3	3.0 ± 0.1	2.1	3.1 ± 0.4	3.0 ± 0.6	2.6 ± 0.3

Note. Number of tokens is in parentheses. EF = English female, EM = English male.

[a] Participant imaged in the United States using a fast spin-echo technique. [b] Participant imaged in Japan using the SMASH protocol.

Table 2. Mean distances (mm ± *SD*; no *SD* for single tokens) at each measurement level for the 6 Japanese speakers.

Speaker	Level	[i]	[e]	[a]	[o]	[ɯ]
JF1[a]	1	3.9		4.0	3.2	3.9
	2	2.4	1.6	2.1	1.8	1.9
	3	2.7	2.1	2.4	2.4	2.1
JM1[a]	2	2.1	2.7	2.5	2.1	2.2
	3	5.2	6.1	5.1	4.7	4.8
JM2[a]	2	2.1	1.6	1.8	2.4	1.8
	3	1.8	1.8	1.8	2.4	1.8
	4	2.6	2.1	2.1	2.6	2.3
JM3[b]	2	1.8	1.8	2.0	1.8	1.5
	3	2.0	1.8	2.3	2.3	1.5
	4	2.3	1.8	2.6	1.8	2.0
JM4[b]	2	1.6	1.5	1.7	1.0	1.4
	3	1.5	2.6	1.7	1.6	1.6
	4	2.1	1.8	2.5	2.1	2.3
JM5[b]	1	1.3 ± 0.0 (2)	1.4 ± 0.1 (2)	1.3 ± 0.0 (2)	1.3 ± 0.0 (2)	1.3 ± 0.3 (2)
	2	1.2 ± 0.1 (2)	1.8 ± 0.0 (2)	1.7 ± 0.2 (2)	1.3 ± 0.4 (2)	1.0 ± 0.0 (2)
	3	1.8 ± 0.4 (2)	1.9 ± 0.2 (2)	1.8 ± 0.0 (2)	1.7 ± 0.2 (2)	2.0 ± 0.2 (2)

Note. Number of tokens is in parentheses. JF = Japanese female, JM = Japanese male. All participants were imaged in Japan.

[a] Participant imaged using a conventional spin-echo protocol. [b] Participant imaged using the SMASH protocol.

FOV = 26.67 cm. Acquisition time lasted about 115 s using this procedure. This long duration was for a complete volumetric scan, however, so the effective acquisition time for the single midsagittal slice used for measurements in this study was much lower. For the remaining participants, images (10-mm thickness) were collected using a faster SMASH (SiMultaneous Acquisition of Spatial Harmonics) protocol having an acquisition time on the order of 2 s: TR = 10 ms, TE = 3.9 ms, 256 × 256, 1 nex, FOV = 25.81 cm (30.97 for 1 participant). Spatial resolution was approximately 1 mm per pixel in the plane of acquisition.

For head stabilization in Japan, participants rested their heads on a rice pillow, which resists deformation. A Velcro strap across the forehead held the head in place. In the United States, participants used a foam pillow, and although they were theoretically capable of moving their heads, the two parasagittal slices obtained in these scans were used to verify that no out-of-plane movement had occurred.

Analysis

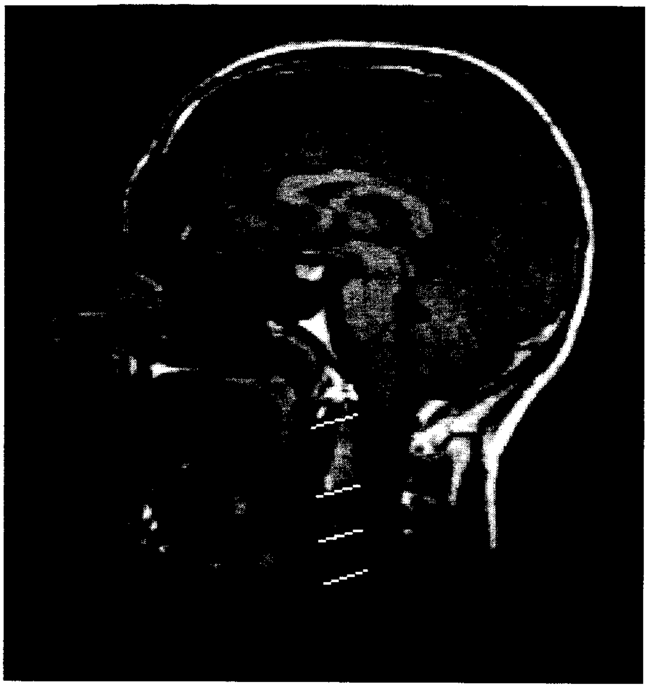
Posterior pharyngeal wall distances were measured from the MR images on a Power Macintosh G4 computer using the public domain NIH Image v1.62 program (available from the U.S. National Institutes of Health at http://rsb.info.nih.gov/nih-image/).

Because our imaging parameters precluded the ability to view the pharyngeal constrictors, and given the differences in vocal tract length among participants, we identified anatomical landmarks across various speakers similar to those used in the swallowing literature (cf. Logemann et al., 2000). Distances were measured between the anterior portion of the vertebral body and pharyngeal wall at four levels whenever possible (see Figure 1). The first measurement was on a line traversing the junction between the dens and body of the second cervical vertebra (C2). The next three measurements were on a line at the inferior borders of the bodies of C2, C3, and C4. All four measurements were possible for only 1 speaker, EM1. For many of the other speakers, the relevant landmarks for the first measurement could not be reliably identified on the images. The fourth measurement point was not applicable in female participants because of their shorter vocal tracts (i.e., more superior larynx position). The individual data for the English and Japanese speakers are shown in Tables 1 and 2, respectively.

Results

Changes in posterior pharyngeal wall position were minimal, with no pattern of differences emerging across

Figure 1. MR image of participant EM1 vocalizing [ɑ] using a fast spin-echo protocol. The four white lines along the anterior portions of the cervical vertebrae indicate the measurement levels.



vowels in either English or Japanese. For both the English and Japanese speakers (Figures 2 and 3, respectively), the variation in measured distances at Levels 2, 3, and 4 shows a similar pattern; the distances mostly lie within a 1-mm band, which is also the extent of the spatial resolution of the MR images. (The distances for English are slightly larger at the first measurement level; however, these data are from only 1 speaker.) The differences in distances for nine English vowels ([ʌ] and [ɔ] were excluded because there were too few data points) were examined in three separate analyses of variance: measurements for Levels 2, 3, and 4; measurements for Level 2 only; and measurements for Level 3 only. There were no significant differences.

We examined the data more closely to determine whether vowel-specific patterns of pharyngeal wall position exist. One likely contributing factor would be pharyngeal width during production of the vowel. That is, a vowel with little or no pharyngeal constriction, such as [i] or [u], may show less posterior pharyngeal wall movement than a vowel with a narrow pharyngeal constriction, such as [ɑ]. However, this does not appear to be the case with the English speakers (Figure 2). Likewise, the Japanese speakers do not exhibit consistent vowel-related trends (Figure 3).

A second possible pattern of pharyngeal wall position among the English vowels might be a difference

Figure 2. Distance from the anterior portion of the vertebral body to the pharyngeal wall averaged over 7 English speakers for each measurement level. Vowels are ordered by increasing pharyngeal constriction (Whalen, Kang, Magen, Fulbright, & Gore, 1999).

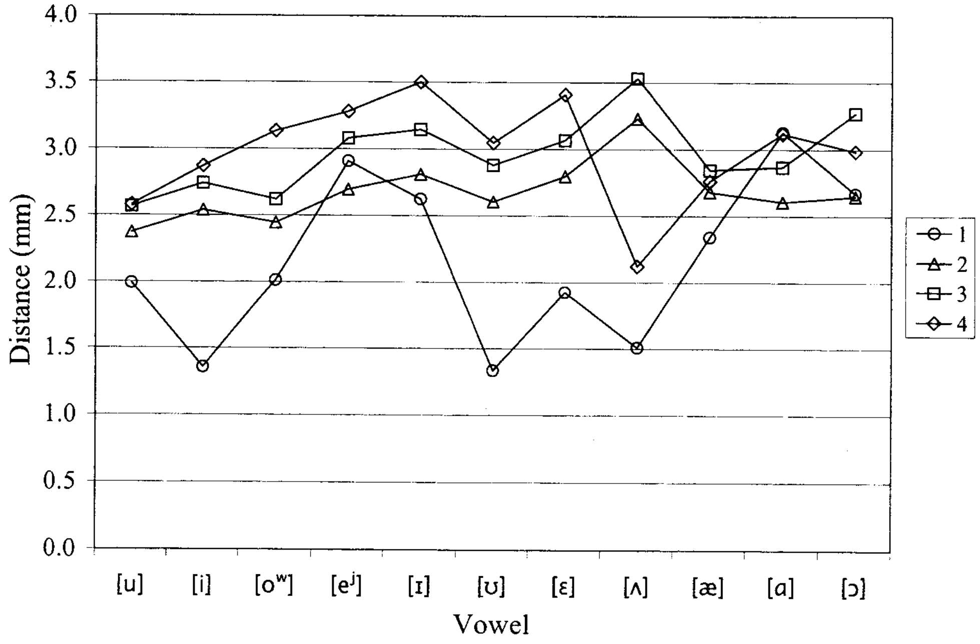
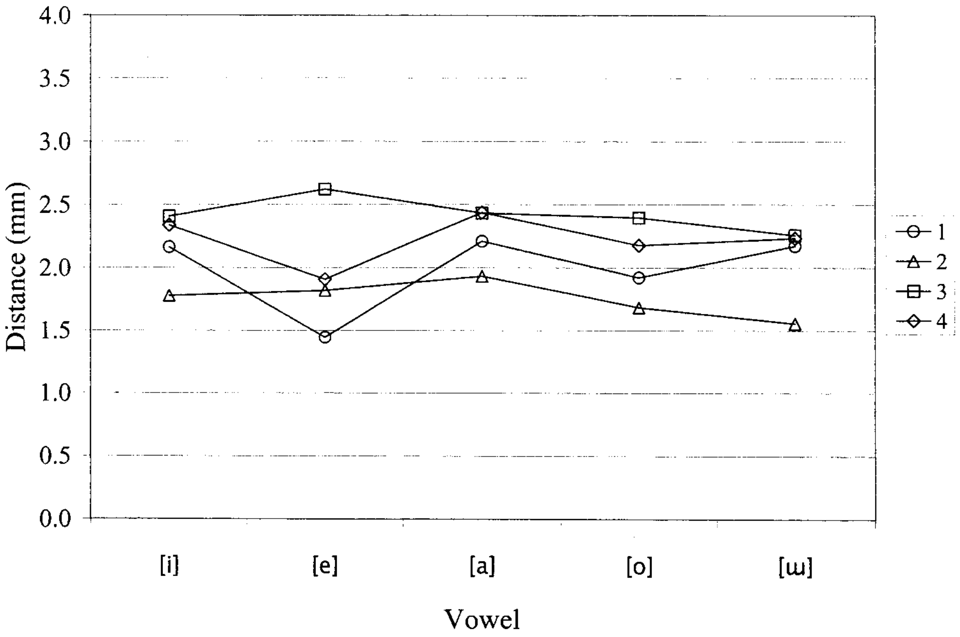


Figure 3. Distance from the anterior portion of the vertebral body to the pharyngeal wall averaged over 6 Japanese speakers for each measurement level. Vowels are ordered as in Figure 2.



between tense and lax vowels: [i ɪ], [eʲ ɛ], [u ʊ]. The production of the tense/lax distinction involves differences in tongue root position for some English speakers (Ladefoged, DeClerk, Lindau, & Papçun, 1972; Perkell, 1971; Tiede, 1996), possibly affecting the posterior pharyngeal wall. Examining the data from the three pairs, it appears that the average distance for the lax member of the pair, 3.0 mm, is slightly greater than that for the tense member, 2.7 mm; however, these differences are far below the resolution of measurement. A third possible pattern in the data might be a difference among vowels of varying heights. Again, these differences are quite small and inconsistent: high vowels show a mean of 2.8 mm; mid vowels, 2.9 mm; low vowels, 2.8 mm.

Experiment 2: X Ray

The formants of vowels depend greatly on their consonantal environment (Hillenbrand, Clark, & Nearey, 2001; Lindblom & Studdert-Kennedy, 1967; Öhman, 1965; Stevens & House, 1963), so the position of the pharynx could conceivably depend on whether there are consonants in the utterance or not.

Furthermore, the MRI data were, by necessity, collected with the participants in a supine position; therefore, the effects of gravity may be a factor in the interpretation of this data. For this reason we supplemented the MR data with cineradiographic data from a participant in a sitting position.

Method

Participant

The participant, EM6, was a 38-year-old male native speaker of Canadian English from Toronto with no known speech, hearing, or neurological problems.

Stimuli

The stimuli were continuous speech of two types: isolated nonsense words and two sentences. The nonsense words were single repetitions of 21 bisyllables of the form hVCV, with one exception, [iɑ]. The following 8 vowels were measured from the nonsense words at Levels 2, 3, and 4 (number of tokens are noted in parentheses): [i] (9), [ɪ] (1), [ɛ] (7), [æ] (1), [ɑ] (7), [ʊ] (1), [u] (1), and [ə] (6). The sentences were, "Why did Ken set the soggy net on top of the deck?" and "I have put blood on her two clean yellow shoes," yielding 24 vowel and 35 consonant measurements.

Apparatus

X-ray films were shot at the cineradiographic facility of the Wenner-Gren Research Laboratory at Nortull's Hospital, Stockholm, Sweden. The films were recorded on high-speed 35-mm film at 45 frames per second. Further details of the original X-ray filming procedures are available elsewhere (Perkell, 1969). The present study used the laser disc copy of the films (Munhall, Vatikiotis-Bateson, & Tohkura, 1995).

Analysis

The X-ray video was captured from laser disc onto a Power Macintosh G4 as a QuickTime movie using MyTV (Eskape Labs, Pleasanton, CA) and Adobe Premiere v6.0 (Adobe Systems, San Jose, CA). Individual frames were saved as PICT files (Figure 4). Lead pellets spaced one centimeter apart in the original films were used for measurement calibration. For each phoneme articulation, the video frame corresponding to the midpoint of the vowel was selected (i.e., the frame at which maximal opening is achieved). Measurements were made from this frame using the same measurement points as for the MR images. The apparent slope is downward in the X-ray images when compared to the MR images, a result of the measurements being referenced to the anatomical landmarks chosen.

Results

The X-ray results, like the MRI results, show relatively little movement of the posterior pharyngeal wall, and the movement is not related to any of the variables examined. As shown in Figure 5, mean vowel distances were within a 1.6-mm range for the second measurement point, within a 1.3-mm range for the third measurement point, and within a 1.2-mm range for the fourth measurement point. A comparison of Figure 5 with Figures 2 and 3 shows that although there is slightly more movement indicated by the X-ray data, the difference is small; therefore, it appears that there is no difference between sustained productions and dynamically produced speech.

To examine possible differences between two dynamic speech contexts (isolated words vs. sentences) and between vowels and consonants as classes of speech sounds, the distances at the approximate midpoint of each phoneme in isolated words as well as in the two sentences produced by EM6 are plotted in Figure 6. As the figure indicates, the range of distances for the individual tokens is virtually identical for three categories of productions: vowels in isolated words, vowels in sentences, and consonants in sentences.

Discussion

Using two imaging techniques, MR and X ray, we examined posterior pharyngeal wall position in order to determine whether it can be eliminated as a variable in vocal tract modeling. MR images of 7 English speakers (2 female, 5 male) and 6 Japanese speakers (1 female, 5 male) and X-ray images of 1 male English speaker were measured at as many as four points from the upper oropharynx to the upper laryngopharynx. Measurement

Figure 4. A single frame from an X-ray film of EM6 during the midpoint of the vowel [ɑ] in the isolated word context. As in Figure 1, the white lines indicate the measurement levels. Landmarks for the first measurement level cannot be resolved.

Figure 5. Distance from the anterior portion of the vertebral body to the pharyngeal wall for participant EM6 (X ray) averaged over tokens for each vowel at Measurement Levels 2, 3, and 4. Vowels are ordered as in Figure 2.

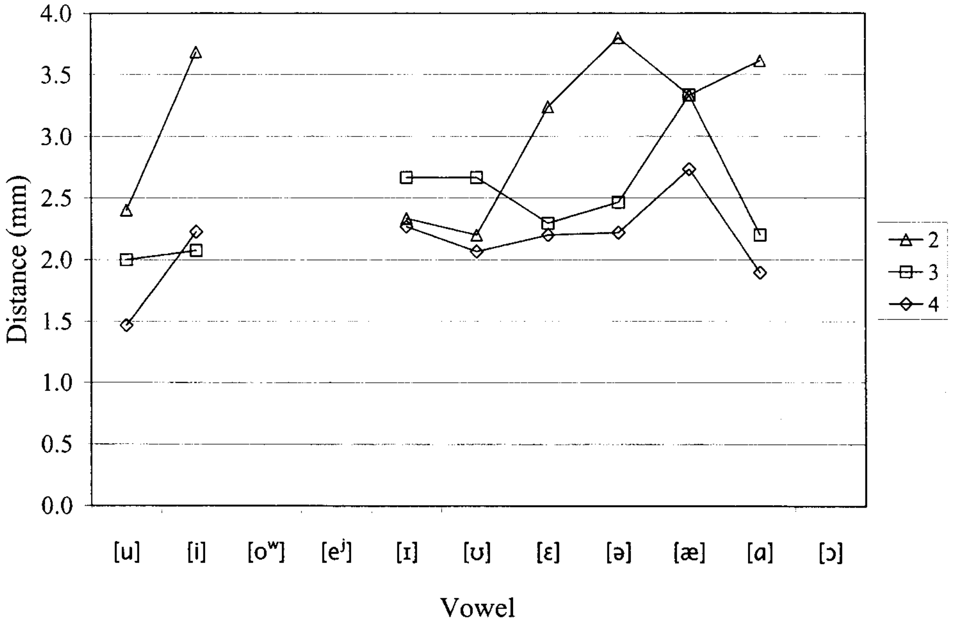
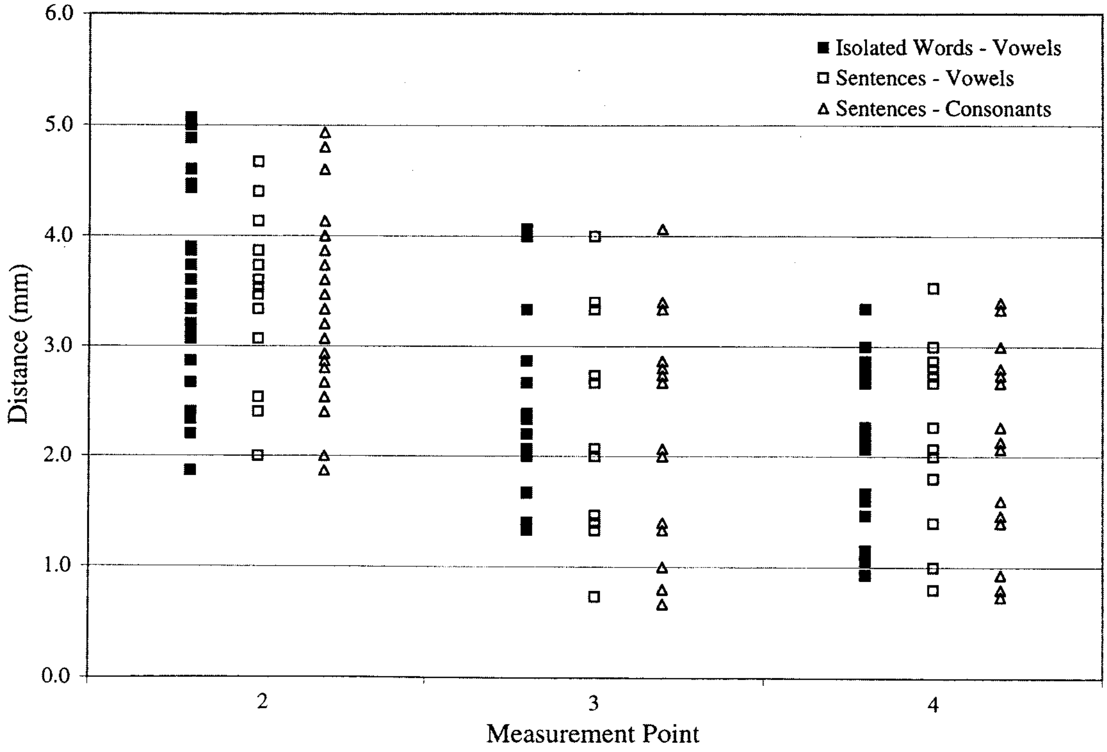


Figure 6. Distance from the anterior portion of the vertebral body to the pharyngeal wall at all measured points for EM6 for dynamically produced speech contexts. Each measurement is taken from the video frame corresponding to the approximate midpoint of phoneme articulation.



in the present study indicates very little movement of the posterior pharyngeal wall during speech, and the movement that is seen is not correlated with vowel differences (e.g., tenseness versus laxness, pharyngeal width, vowel height), vowels and consonants as classes of speech sounds, method of production (sustained or dynamically produced speech), or production context (isolated words or sentences). Participants studied were normal speakers; therefore, results are not necessarily applicable to disordered populations.

The degree of variation in posterior pharyngeal wall position in speech can be compared to findings concerning this structure in the process of swallowing. As discussed above, the posterior pharyngeal wall shows substantial movement during swallowing. The videoradiographic study of swallowing performed by Palmer et al. (1988), examining a measurement point about 1 cm above our first measurement level, showed a range of movement from 4 to 7 mm; examining a measurement point roughly the same as our second measurement point, the study showed a range of movement of 4 mm. The videofluoroscopic study of swallowing in men of two age groups, using a measurement point comparable to our second measurement point, showed structural movement of the posterior pharyngeal wall of 7 mm for younger men and 5.8 mm for older men in the anterior direction (Logemann et al., 2000). Thus, it appears that although the posterior pharyngeal wall can move as much as 7 mm during swallowing, very little movement is seen during speech.

Neither of the languages used in the present study employ pharyngeal distinctions. One language that contrasts vowels on the basis of pharyngeal size, the West African language Akan, does not show appreciable posterior pharyngeal wall movement, except for a small amount of movement in the lower pharynx related to laryngeal height (Lindau, 1975). It is possible that other languages, especially those with pharyngeal consonants, such as Arabic or the Athabascan language Tlingit, may show a different pattern.

The possibility exists that there is a difference between the position of the posterior pharyngeal wall at rest and during speech. We did not fully investigate this possibility because our focus was on speech. Measurements from rest data available for 1 English participant and 1 Japanese participant showed no difference for the English participant, but suggested a small difference for the Japanese participant. However, due to the limited availability of rest data, these results must be considered inconclusive.

The two imaging methodologies used in this study each have advantages, but results from each need to be interpreted with caution. For MR, in-plane resolution (i.e., anterior-posterior and superior-inferior dimensions on the sagittal images) was 1 mm for all three imaging protocols, but they differed in slice thickness (3, 5, and 10 mm). Although these thicknesses are better than the extreme situation presented by X-ray imaging in which the entire thickness of the head is reduced to a single two-dimensional image, a 10-mm-thick sagittal averaging may produce misleading results. For example, constriction of the posterior portion of the lateral walls (adjacent to the posterior pharyngeal wall) could occur to a sufficient degree to mimic posterior wall movement on the image. In our study, however, because significant posterior wall movement was not evident, this was not an issue.

Participants were supine for MR imaging but upright for X-ray studies. The similarity in the results from the two methodologies and the fact that they involved supine and upright positions provides tentative support for the conclusion that gravitational influences were negligible.

Present results indicated that the small amount of posterior pharyngeal wall movement during speech is not related to phonetic category or production method. This finding should justify the elimination of the posterior pharyngeal wall as a variable in estimation of the vocal tract cavity in many languages. Consequently, measurements of the anterior pharyngeal wall alone, obtainable through ultrasound imaging, can be used for reconstructing the entire vocal tract. Ultrasound imaging affords noninvasive data collection and real-time imaging. Images of the tongue surface and hard palate can be aligned to provide a nearly complete midsagittal outline of the upper speech airway (Gick, in press). By matching these ultrasound images to independently obtained MR images, the pharyngeal region can be modeled, giving us an unprecedented view of the vocal tract in real time.

Acknowledgments

This research was supported by NIH Grant DC-02717. The first author (HSM) was partially supported by an award from the Rhode Island College Faculty Research Council. Japanese MRI data were collected with the support of the ATR Human Information Processing Laboratories. We are grateful to Khalil Iskarous for technical assistance at Haskins Laboratories and Matthew Richardson for helpful discussions on statistics. Portions of this research were presented at the 142nd Meeting of the Acoustical Society of America, December 2001, in Fort Lauderdale, FL.

References

Bell-Berti, F. (1975). Control of pharyngeal cavity size for English voiced and voiceless stops. *Journal of the Acoustical Society of America*, *57*, 456–461.

Calnan, J. (1954). The error of Gustaf Passavant. *Plastic and Reconstructive Surgery*, *13*, 275–289.

Carmody, F. (1941). An X-ray study of pharyngeal articulation. *University of California Publications in Modern Philology, 21*, 377–384.

Cunningham, E. T., Jr., Donner, M. W., Jones, B., & Point, S. M. (1991). Anatomical and physiological overview. In B. Jones & M. W. Donner (Eds.), *Normal and abnormal swallowing: Imaging in diagnosis and therapy* (pp. 7–32). New York: Springer–Verlag.

Gick, B. (in press). The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association.*

Hagerty, R., & Hill, M. (1960). Pharyngeal wall and palatal movement in postoperative cleft palates and normal palates. *Journal of Speech and Hearing Research, 3*, 59–66.

Hagerty, R., Hill, M., Pettit, H., & Kane, J. (1958). Posterior pharyngeal wall movement in normals. *Journal of Speech and Hearing Research, 1*, 203–210.

Harrington, R. (1944). A study of the mechanism of velopharyngeal closure. *Journal of Speech Disorders, 9*, 325–345.

Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America, 109*, 748–763.

Iglesias, A., Kuehn, D. P., & Morris, H. L. (1980). Simultaneous assessment of pharyngeal wall and velar displacement for selected speech sounds. *Journal of Speech and Hearing Research, 23*, 429–446.

Kent, R. D., & Moll, K. L. (1969). Vocal-tract characteristics of the stop cognates. *Journal of the Acoustical Society of America, 46*, 1549–1555.

Ladefoged, P., DeClerk, J., Lindau, M., & Papçun, G. (1972). An auditory-motor theory of speech production. *Working Papers in Phonetics, 22*, 48–75. Los Angeles: University of California.

Lindau, M. (1975). [Features] for vowels. *Working Papers in Phonetics, 30.* Los Angeles: University of California.

Lindblom, B., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America, 42*, 830–843.

Logemann, J. A., Pauloski, B. R., Rademaker, A. W., Colangelo, L. A., Kahrilas, P. J., & Smith, C. H. (2000). Temporal and biomechanical characteristics of oropharyngeal swallow in younger and older men. *Journal of Speech, Language, and Hearing Research, 43*, 1264–1274.

Minifie, F. D., Abbs, J. H., Tarlow, A., & Kwaterski, M. (1974). EMG activity within the pharynx during speech production. *Journal of Speech and Hearing Research, 17*, 497–504.

Minifie, F. D., Hixon, T. J., Kelsey, C. A., & Woodhouse, R. J. (1970). Lateral pharyngeal wall movement during speech production. *Journal of Speech and Hearing Research, 13*, 584–594.

Moll, K. L. (1960). Cinefluorographic techniques in speech research. *Journal of Speech and Hearing Research, 3*, 227–241.

Moll, K. L. (1962). Velopharyngeal closure in vowels. *Journal of Speech and Hearing Research, 5*, 30–37.

Munhall, K. G., Vatikiotis-Bateson, E., & Tohkura, Y. (1995). *X-ray film database for speech research* [videodisc]. Kyoto, Japan: ATR Laboratories.

Öhman, S. E. G. (1965). X-ray studies of articulatory dynamics. *Proceedings of the 5th International Congress of Phonetic Sciences*, 16–23.

Palmer, J. B., Tanaka, E., & Siebens, A. A. (1988). Motions of the posterior pharyngeal wall in swallowing. *Laryngoscope, 98*, 414–417.

Passavant, G. (1863). *Über die verschliessung des schlundes beim sprechen* [Concerning the pharynx during speech]. Frankfurt am Main: J. D. Sauerländer.

Perkell, J. S. (1969). *Physiology of speech production: Results and implications of a quantitative cineradiographic study.* Cambridge, MA: MIT Press.

Perkell, J. S. (1971). Physiology of speech production: A preliminary study of two suggested revisions of the features specifying vowels. *Quarterly Progress Report, 102*, 123–139. Cambridge, MA: Research Laboratory of Electronics, Massachusetts Institute of Technology.

Perlman, A. L., Luschei, E. S., & DuMond, C. E. (1989). Electrical activity from the superior pharyngeal constrictor during reflexive and nonreflexive tasks. *Journal of Speech and Hearing Research, 32*, 749–754.

Seikel, J. A., King, D. W., & Drumright, B. S. (2000). *Anatomy and physiology for speech, language and hearing* (2nd ed.). San Diego, CA: Singular.

Shiller, D. M., Ostry, D. J., & Gribble, P. L. (1999). Effects of gravitational load on jaw movements in speech. *Journal of Neuroscience, 19*, 9073–9080.

Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research, 6*, 111–128.

Tiede, M. K. (1996). An MRI-based study of pharyngeal volume contrasts in Akan and English. *Journal of Phonetics, 24*, 399–421.

Tiede, M. K., Masaki, S., & Vatikiotis-Bateson, E. (2000). Contrasts in speech articulation observed in sitting and supine conditions. *Proceedings of the 5th Seminar on Speech Production: Models and Data*, 25–28.

Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America, 73*, 1322–1336.

Whalen, D. H., Kang, A. M., Magen, H., Fulbright, R. K., & Gore, J. C. (1999). Predicting pharynx shape from tongue position during vowel production. *Journal of Speech, Language, and Hearing Research, 42*, 592–603.

Zagzebski, J. A. (1975). Ultrasonic measurement of lateral pharyngeal wall motion at two levels in the vocal tract. *Journal of Speech and Hearing Research, 18*, 308–318.

Zemlin, W. R. (1998). *Speech and hearing science: Anatomy and physiology* (4th ed.). Englewood Cliffs, NJ: Prentice–Hall.

Received June 13, 2002

Accepted August 7, 2002

DOI: 10.1044/1092-4388(2003/019)

Contact author: Harriet S. Magen, PhD, Haskins Laboratories, 270 Crown St., New Haven, CT 06511. E-mail: hmagen@ric.edu