Title: Using active shape modelling based on MRI to study morphological and pitch-related functional changes affecting vocal structures and the airway

Abstract

Objective

The shape of the vocal tract and associated structures (e.g. tongue and velum) is complicated and varies according to development and function. This variability challenges interpretation of voice experiments. Quantifying differences between shapes and understanding how vocal structures move in relation to each other is difficult using traditional linear and angle measurements. With statistical shape models, shape can be characterized in terms of independent modes of variation. Here, we build an active shape model (ASM) to assess morphological and pitch-related functional changes affecting vocal structures and the airway.

Method

Using a cross-sectional study design, we obtained 6 midsagittal MR images from 10 healthy adults (5 males and five females) at rest, while breathing out, and while listening to and humming low and high notes. Eighty landmark points were chosen to define the shape of interest and an ASM was built using these (60) images. Principal component analysis was used to identify independent modes of variation and statistical analysis was performed using one-way repeated-measures ANOVA.

Results

Twenty modes of variation were identified with modes 1 and 2 accounting for half the total variance. Both modes 1 and 9 were significantly associated with humming low and high notes ($P < 0.001$) and showed coordinated changes affecting the cervical spine, vocal structures and airway. Mode 2 highlighted wide structural variations between subjects.

Conclusion

This study highlights the potential of active shape modelling to advance understanding of factors underlying morphological and pitch-related functional variations affecting vocal structures and the airway in health and disease.

Keywords

active shape model; active appearance model; MRI; vocal tract; pitch; humming; cervical spine; posture

Introduction

Speech and singing are complex activities requiring rapid and finely coordinated movements of muscles responsible for articulation, phonation and respiration. (1) Until the 1980s, available methods, such as ultrasound, electropalatography, and nasoendoscopy, meant that only part of the vocal apparatus could be examined at any one time. However, with the introduction of magnetic resonance imaging (MRI) in voice research it became possible to investigate the soft tissue outline of the entire vocal tract (glottis to lips) in three dimensions. (2) Since then, advances in technology and reduction in image acquisition time mean that MRI can now be used to observe vocal function in real time. (3) Nevertheless, despite these significant advances, we do not yet have a comprehensive understanding of factors underlying the wide range of structural variation between individuals and functional variations within and between individuals. (4,5) This is important because uncovering factors responsible for such variability could lead to new insights and, therefore, testable hypotheses concerning fundamental questions in voice science (6): for example, what mechanisms underlie the “singing formant” (7) and the rise and fall of the larynx with changes of voice pitch? (8,9)

Traditionally, the focus of MRI in voice science is restricted to the investigation of changing dimensions of the vocal tract and vocal structures, and changing relationships between articulators such as the lips, jaws, tongue, and soft palate (velum). (1,10) In previous studies, it was suggested that this focus was too narrow because such an approach neglects to account for the fact that all vocal structures have direct and/or indirect attachments to the skeletal frame; the skull, cervical spine, sternum and scapula. (11,12) Instead, it was argued that by considering vocal structures within the context of their wider relationships it might be possible to reach a better understanding of mechanisms underlying coordinated adjustments responsible for goal-related activity within the vocal system. Using a method that combined MRI’s superior soft-tissue definition with bony reference points used in cephalometry (lateral x-ray), a protocol was designed to allow investigation of morphological and dynamic functional relationships between vocal structures within the context of their anatomical connections.

In the first study, with subjects at rest, widespread and significant correlations were observed between variables relating the larynx, hyoid, epiglottis, velum and airway to the cranial base, craniofacial skeleton, sternum, and cervical spine. (11) These included previously unreported correlations (e.g. between the width of the laryngeal tube opening and craniocervical posture). In the second study, images were acquired while subjects hummed low and high notes while maintaining a stable posture. (12) Significant differences were found between low- and high-note conditions in 6 of 22 measures in addition to widespread significant pitch-related correlations between variables. Specifically, compared with humming a low note, humming a high note was associated with a rise of the larynx and hyoid in relation to the cranial base, increased angles between the cranial base and cervical spine, and increased C3-menton and sternum-hyoid distances. These results demonstrated the presence of coordinated pitch-dependent adjustments during voice production that may be missed or mistakenly attributed to articulatory or postural changes, particularly if vocal structures are investigated without taking their wider structural relationships into account.

In both these studies, significant correlations between variables were reported in correlation tables. It is difficult to gain a full appreciation of underlying patterns of adjustments that accompany both developmental and functional changes within the head and neck from complicated arrays of numbers displayed in a correlation table. However, by using data obtained in these studies to build a statistical shape model it is possible to present these findings in a more accessible and informative *visual* format.

Background

The vocal tract and its closely associated structures such as the larynx, hyoid, epiglottis, tongue, jaw and velum vary not only in shape and size between individuals, significant variation also occurs within individuals during voice production. The length of the vocal tract, for example, varies according to age, sex and size (4), and the shape of the airway can vary according to changing posture (13) and changing positions of articulators. (1) These variations in the overall shape of vocal structures and the airway can easily be seen in midsagittal MR images of the head and neck. However, statistical comparison of images requires a valid method of quantifying such variation. In comparison with a simple shape such as a rectangle, MR images and changing relationships between structures they represent are complex and difficult to represent mathematically. An active shape model (ASM) is a statistical model that can account for such natural variation. (14-17)

Active shape modelling is a well-established image processing technique that can be used in situations where, as here, objects of interest can be clearly defined and a representative set of examples is available. Since description of the first flexible deformable model to allow for such natural variability (18), statistical models have earned their place as a “systematic and effective paradigm for the interpretation of complex images.” (16) They have wide application in a growing list of medical disciplines. This includes modelling of arthritic and osteoporotic hips (19), vertebrae (20), facial appearance (15), the heart (15), brain ventricles (15) and, more recently, speech production. (21-24) In an active shape model, shape is represented mathematically (by recording coordinates of points) and incorporated into a flexible template. In addition to shape, an active appearance model also represents other shapes, surrounding structures and boundaries by registering their grey-level or texture appearance (the pattern of pixel intensities which varies according to tissue type). As shape and texture are often correlated, combining information about both aspects means that a more informative model can be obtained. (24)

In this study, the shape of interest includes vocal structures, the airway, and bony landmarks that allow these structures to be studied within the context of their wider relationships: Vocal structures are not isolated. They are anatomically and functionally linked to surrounding structures (e.g. superior constrictor connection with tongue muscles) (25), and to spatially more distant structures (e.g. connections between velum and larynx via palatopharyngeus). (26) Using statistical shape and appearance models, it is possible to observe and quantify correlations between positions and shapes of local and more distant structures within the image: that is, it is possible to observe how different vocal structures move in relation to each other as the overall shape varies according to development or function. This approach has been termed a ‘top-down’ (global) rather than a ‘bottom-up’ strategy. (14) In the latter instance the focus is on local structures and their relationships; between the diameters of the laryngeal tube opening and the hypopharynx, for example. (27) However, the complexity of MR images and the almost limitless way that structures may vary in relation to each other means that a ‘bottom-up’ approach is necessarily restrictive in what it can reveal. Recently, statistical models of shape and appearance were successfully used to model tongue shape and motion (21) and vocal tract shape (24) during articulation of speech sounds. As far as we are aware, this is the first statistical shape model to represent vocal structures from a global perspective; one that takes into account their wider anatomical relationships within the skeletal frame.

The aims of this study are threefold: 1) to model differences in gross morphological features of the vocal tract and associated structures within the head and neck between subjects at rest 2) to model changes in shape that occur when subjects hum low and high notes and 3) to show how active shape modelling can complement and extend information obtained by using more traditional geometric measurements.

Method

The selection of subjects and a description of the method used to acquire midsagittal MR images were reported earlier. (11,12) In brief, MR images were acquired from a mixed group of singers (including one professional singer) and non-singers (5 males, 5 females aged between 20-47 years with a median of 25 years). Before image acquisition, the low and high notes that could be comfortably hummed while breathing out over 20 seconds were established. For the whole group, these ranged from 98 (G2) to 1047 Hz (C6), where C4 is middle C. Subjects adopted a supine position in the scanner and were instructed to maintain a stable posture at all times (looking straight ahead with lips and teeth together and tongue resting comfortably against the hard palate). Individuals were imaged with the head placed in a Sense-Neurovascular array-16 element coil. Deformable foam wedges were used to make the subject comfortable and restrain the head position. Parasagittal images were obtained with a 3.0 T Achieva MR system (Philips, Best, Holland) using a turbo spin echo pulse sequence with the following parameters: field of view (FOV) 340 mm x 340 mm; a 768 by 768 matrix; repetition time 4106 ms; echo time 100 ms; 6 slices 4.0 mm thick with a gap of 1.0 mm centred on the midsagittal plane. As part of a larger study, six images were acquired from each of the 10 subjects while at rest, while humming low and high notes, while listening to the same low and high notes, and while breathing out over 20 seconds. Only data referring to images acquired while at rest and while humming low and high notes are analyzed in this paper.

Building the Active Shape Model

The model was built using a freely available active appearance modelling tool kit from the University of Manchester (http://www.isbe.man.ac.uk/~bim/software/am_tools_doc/index.html). Examples of MR images used to build the model are shown in Figure 1. Point selection and annotation of MR images was carried out by one of the authors, a clinician with detailed knowledge of vocal anatomy and relevant imaging experience. Steps taken to build the model are illustrated in Figure 2 and summarized below:

Eighty points were chosen to describe the shape of interest (Figure 3). These included a) well-defined ‘landmark’ points, easy to locate on every image and corresponding to particular features such as the tips of the velum, epiglottis and odontoid process, and b) boundary points (equidistant between landmark points) which help define the shape of interest and assist in the visual interpretation of results. The points chosen for this study included all those defined and selected for conventional geometric analysis in earlier studies. (11,12)

A template representing the shape was obtained by manual annotation of the first image. Each point was carefully and precisely placed on the same feature (on this and each subsequent image) and the way in which points were connected was recorded so that the method could model variability effectively. (This process provides a crude template modelled on only one set of points; a model of the shape and ‘texture’ of the image. If used to locate the same shape in a new image, this model would only be able to map on to shapes that are almost identical to itself; i.e. it is a rigid, rather than a flexible, template.) A flexible template, containing all the shape variations present in the data set, was obtained by uploading each of the remaining 59 images in turn. For each image, the software attempts to match the model to the new set of points. Precise matching of the model to the new shape was achieved following careful manual editing of the position of each point. Once matched, the model was updated. (The updated model incorporates the shape represented in the newly uploaded image and, therefore, the ways this shape differs from the original image.) As each set of points was uploaded to the model they were aligned into a common coordinate frame by scaling, rotation and translation (Procrustes analysis) to minimize the variance, in distance, between equivalent points. With the scaling factor removed all the data is stored proportionately rather than absolutely. This means that the effect of subject size on measurements such as vocal tract length is eliminated, allowing the shapes themselves to be compared. Once all the data is incorporated into the model, the software calculates the average position of the points to obtain the mean shape of the chosen structures and its allowable variations (Point Distribution Model): individual modes of variation are calculated using principal component analysis (PCA).

PCA is a powerful tool that is widely used to uncover hidden patterns in data. Using deviations from the mean, it identifies ways in which groups of landmark points tend to move in relation to each other as the shape varies. Each identified pattern of movement (or overall change of shape) represents a statistically independent mode of variation (i.e., a change in shape that occurs independently of other shape changes). When combined, the modes of variation account for 100% of variance in the data set. Mode numbers are ordered according to the amount of variation explained with mode 1 accounting for the largest proportion of variance in shape, and higher mode numbers accounting for progressively smaller proportions of variance. For each mode in the model, the mean and standard deviation (SD) value for the whole MRI data set (60 images) was calculated and scaled to zero mean and unit SD. The score for each mode was then calculated for each image and expressed in terms of how many SDs it lay from the mean referent value (zero) of that mode (i.e. how its shape compares to others in the group). The scores from these modes of variation were used as inputs to the statistical analysis.

Statistical analysis

Statistical analysis was performed using *Sigmastat* (Version 11; Systat Software, Inc., San Jose, CA). One-way repeated-measures analysis of variance (ANOVA) was used to investigate differences between groups, and post hoc ANOVA group comparisons were performed using the Holm-Sidak test with significance set at $P \leq 0.05$. For all tests, $P < 0.05$ was taken to indicate statistical significance. The images were also analyzed using conventional geometric measurements, the results of which were reported earlier. (11,12)

Results

Findings of the Active Shape Model

The first 20 modes accounted for 98% of the total variance (Table 1) but after mode 10 none contained more than 1% of the variance. The first 2 modes accounted for half the total variance. Here we report findings associated with modes 1, 2 and 9. To assist understanding, key anatomical features are illustrated in Figure 4. Mode scores for modes 1 and 9 were significantly different between humming low and high notes ($P < 0.001$). Mode 1 scores changed from 0.18 to -0.44 and mode 9 from -0.29 to 1.06 on changing pitch from low to high. Figure 5 shows the shapes described by varying modes 1 and 9 by +2 to -2 SDs about the mean shape of all 60 images. The same information is available as more informative and visually compelling video demonstrations by clicking on Figure 5 in the on-line version of the Journal. Mode 1 is associated with coordinated changes affecting the cervical spine, vocal structures, and airway (nasopharynx to hypopharynx). Specifically, increasing kyphosis of the cervical spine is associated with shortening of the airway; a rise of the larynx (upper C6 to upper C4), hyoid (lower C3 to top of C5), epiglottis tip (bottom of C3 to lower C2), and velar tip (bottom of C2 to upper C2); increasing distance between the sternum and larynx; and, decreasing distance between the larynx and hyoid. Conversely, increasing lordosis of the cervical spine is associated with lengthening of the airway; lowering of the larynx, hyoid, and epiglottis and velar tips; decreasing distance between the sternum and larynx; and, increasing distance between the larynx and hyoid. Changes of airway length are also associated with changes affecting the midsagittal shape of the nasopharyngeal and hypopharyngeal cavities: increasing airway length appears to be associated with a larger nasopharyngeal cavity and a longer and narrower hypopharyngeal cavity, whereas reductions in length appear to be associated with smaller nasopharyngeal dimensions and a shorter, wider hypopharyngeal cavity. The size of the oropharyngeal cavity did not appear to change.

Mode 9, although accounting for only 1.82% of the total variance within the data set, is also associated with coordinated changes affecting craniofacial and cervical structures. In contrast with mode 1, cervical changes, although present (lordosis to kyphosis), are slight. Additionally, in relation to the cervical spine, the heights of the larynx, hyoid, epiglottis and velum remain unchanged. However, whereas mode 1 is associated with changing airway *length*, mode 9 is associated with changing airway *width*: That is, as the cervical spine moves towards kyphosis, the velopharyngeal opening (VPO) and oropharyngeal airway become narrower and the hypopharyngeal opening becomes wider. Of particular interest in this mode is the finding that there appears to be a reciprocal relationship between midsagittal dimensions of the VPO and the shape of the geniohyoid muscle.

The second mode accounts for 20.4% of the variance. Although no statistically significant differences were found between the six conditions, this mode is important because it highlights natural variations in the overall head-neck shapes within this group of 10 subjects. Mode 2 is associated with variations in the shape of craniofacial structures in relation to the cervical spine. The effect of altering this mode of variation by +2 to -2 SDs about the mean shape of all 60 images is seen in Figure 6 and also in a more informative video demonstration by clicking on Figure 6 in the on-line version of the Journal. As the cervical spine moves from kyphosis to lordosis, the distance between the larynx and sternum increases and the distance between the larynx and menton decreases. Narrowing of the VPO is accompanied by a rise of the velum and a reduction of the angle between the hard and soft palate. Rotation of the hyoid is accompanied by lowering of the epiglottis and posterior displacement of the tongue. The height of the larynx remains unchanged and there appears to be little change in the relationship between the upper cervical spine and the alignment of the base of the skull.

Discussion

Speech and singing require finely coordinated movements of muscles responsible for articulation, phonation and respiration. (1) We lack a full understanding of the mechanisms responsible for such coordinated activity. (28,29) In this study, midsagittal MR images of the head and neck were used to build an active shape model to investigate morphological differences of vocal and associated structures within the head and neck, and to investigate changes in the shape of these structures when subjects hummed low and high notes. Our results highlight the potential of ASM to significantly improve our understanding of coordinated mechanisms that underlie vocal behavior. Not only can ASM be used to identify and distinguish between structural and functional changes in the shape of vocal structures and the airway, it can also show how vocal structures move together as overall shape varies according to development or function, thus highlighting a key advantage of statistical shape modelling over conventional geometric analysis:

The results of geometric analysis were reported earlier. (11,12) Although geometric analysis and active shape modelling both showed the switch from low- to high-note humming to be accompanied by significant changes in vertical and horizontal dimensions, use of active shape modelling also permitted the discovery of distinct modes of variation that appear to underlie these changes: there is not a 1:1 relationship between functional movements and modes of variation because goal-related movements may require the simultaneous recruitment of two or more modes of variation. By uncovering previously hidden patterns of movement underlying goal-related vocal activity, active shape modelling complements and extends results obtained from conventional geometric analysis.

The model created 20 modes of variation with the first 2 modes accounting for half the total variance within the data set. The modes are in order of decreasing variance, reflecting a reducing measure of global changes in morphology. Two modes of variation (modes 1 and 9) were significantly associated with humming low and high notes. Each shows a different pattern of coordinated activity affecting the cervical spine, vocal structures and the airway. It can be seen from Table 1 that mode 9 accounts for only 1.82% of total variance. However, it is important to appreciate that mode number does not necessarily equate with clinical importance. This is particularly true in voice production studies because local changes of shape can lead to significant effects even if overall shape changes are small: the effect of the size of the VPO on acoustic output, for example. (30) The findings associated with modes 1, 9 and 2 are discussed in turn below.

Mode 1

In mode 1, as the humming pitch changes from low to high, the cervical spine moves from lordosis towards kyphosis, the airway becomes shorter, vocal structures (larynx, hyoid, epiglottis and velum) rise together in relation to the cervical spine, the distance between the sternum and larynx increases and the distance between the larynx and hyoid decreases. Although pitch-related changes affecting the larynx were not altogether surprising, those affecting the alignment of the cervical spine were not anticipated. Below, we consider a number of factors that may have contributed to these changes, beginning with pitch-related changes affecting larynx height.

Pitch change and larynx height

Although the rise and fall of the larynx with pitch is long recognized (e.g. Bérard 1755) (31), the mechanisms underlying this close association are still unclear. (9,10,32) Numerous (direct and indirect) muscular, membranous, and ligamentous attachments to vocal structures and the skeletal frame, functional changes, postural adjustments, and gravitational influences mean that the height of the larynx at any one time depends upon the net force acting on it at that particular moment. These changes are supported by an immensely rich reflex network which serves to integrate the primary demands of the respiratory system with other concurrent task-related activities involving the same structures (e.g. vocalization). (33) In this study, subjects adopted a supine position in the scanner. Compared with upright subjects, the larynx tends to be higher in the supine position, thought to be due to the lack of gravitational pull of the respiratory apparatus (Hixon 1987). (10) Lung volume can also influence laryngeal height with higher lung volumes associated with lower laryngeal positions (Iwarsson 1998). (10) This is important because the act of humming is inevitably accompanied by reducing lung volumes and, therefore, a tendency towards higher laryngeal positions.

In this study, a stable posture, supine position, and sustained phonation were common to both low- and high-note humming conditions, suggesting that pitch-related changes may also contribute to the changes of larynx height observed here. Support for this view is found in the results of an almost identical study which investigated changes when subjects hummed notes at each end of their range whilst adopting an upright position (lateral x-ray), where humming high notes was "undoubtedly" accompanied by upward movement of the "larynx as a whole". (34) More recently, Yanagisawa et al. (1991) observed the rise of the larynx with pitch to be associated with contraction of the pharyngeal walls and commented that pharyngeal constrictor contraction could result in a "dorsocranial pull" (8); observations supported by knowledge of pharyngeal constrictor attachments to the base of the skull and thyroid cartilage (26), reports of pitch-related activity involving pharyngeal muscles (superior and inferior constrictor muscles and palatopharyngeus) (32), and findings of phonation-induced contractile reflexes involving the inferior pharyngeal constrictor and upper esophagus. (35) We suggest that, together, the results of this and earlier studies point to the presence of active pitch-related adjustments that can augment or override biomechanical restraints such as those imposed by respiratory demands and changes of posture or position. This possibility is particularly interesting given suggestions that alternative pitch mechanisms may account for the ability of alaryngeal speakers to convey prosody successfully. (36)

Pitch change and cervical alignment

The extent of cervical involvement in the first mode was unexpected, striking and counterintuitive. Beyond suggestions that regional changes of cervical spine shape might contribute to fine adjustment of fundamental pitch (37), cervical input has no place in traditional theories of pitch production. (38) Cervical changes have been reported in professional singers but they have been attributed to jaw opening (39) or the adoption of a more forward head posture. (40) Compared with the lumbar spine, which tends to maintain its intrinsic shape in upright and supine positions (20), factors affecting the shape of the cervical spine have received little attention. Kitamura et al. (2005) noted that the cervical spine and posterior pharyngeal wall appeared to be retracted backwards in the supine posture and suggested that fixed head position could change the orientation of the head relative to the axis of the body. (41) However, like larynx height, as the same position was adopted in both humming conditions, it is possible that the switch from low-to high-note humming may also have contributed to the cervical changes observed in this study. The “tripartite muscle arrangement” of longus colli (deep cervical flexor) supports this suggestion because, with its origins and attachments confined to the cervical vertebrae, it is well-placed to support functional changes in the shape of the cervical spine. (42)

The changes of cervical shape with pitch led us to ask 1) whether cervical muscles have a greater role in pitch production than previously thought and 2) given the extent of the rise of vocal structures in relation to the cervical spine, whether there is a neural connection that could link, or synchronize, cervical with pharyngeal activity. A greater role for cervical muscles is supported by reports of: increasing distance between the larynx and cervical spine at the higher pitches (32); forward movement of the posterior pharyngeal wall during voice production and pitch-related activity in neck flexors (e.g. longus capitis) (43); and, “markedly elevated” activity in neck muscles when singing the highest pitches. (44) Involvement of pharyngeal muscles in pitch-related activity finds support in experiments showing a rise in pitch to be associated with increased activity in pharyngeal constrictor muscles and palatopharyngeus. (32) Examination of underlying neural connections shows that: nerves supplying the upper occipital, cervical and geniohyoid muscles have a common origin, the first cervical spinal nerve (C1); there is an overlap of the origin of nerves supplying both the deep cervical flexor muscles (e.g. longus colli) and the supra- and infra-hyoid (strap) muscles (C1-C5 and C1-C3 respectively); and, a neural connection “of undetermined function” links the vagus nerve (supplying pharyngeal muscles) to the cervical plexus at the level of C1. (45) The common origin of the nerve supply to cervical and strap muscles led us to ask whether functional synergies might exist not only between cervical and pharyngeal muscles but also between muscles that lie in front of and behind the airway. Such synergy, if confirmed, would have important implications for voice science. We know, for example, that strap muscles have a role in pitch production, particularly during the production of low and high notes, but the nature of this role is unclear. (46) Our results, together with knowledge of underlying neuroanatomical connections lead us to suggest that recruitment of strap and cervical muscles occurs as part of more widespread coordinated activity during pitch production.

The presence of synergy between cervical and strap muscles has important clinical implications. Muscle tension dysphonia (MTD), for example, is characterized by excessive tension in extrinsic or (para) laryngeal musculature. (47) Primary MTD, where dysphonia occurs in the absence of organic vocal pathology, affects up to 40% of those attending voice clinics. Multiple factors are thought to underlie its development but these are not fully understood. Here, we ask whether postural neck muscles, like laryngeal muscles, could be considered as falling into two groups, with postural input from superficial neck flexors (e.g., sternocleidomastoid (SCM)) influencing the functional efficiency of deep cervical flexors (longus colli and longus capitis) and, therefore, the degree of pitch-related cervical spine movement; a view supported by evidence of synchronous activity between longus colli and SCM. (48) Overall, however, rather than synergies between these particular muscles or muscle groups it is, perhaps, more profitable to view the potential for synergy between different muscles in the head-neck region as being more widespread than previously thought, with activity in any one muscle or muscle group varying according to task demands at the time.

Mode 9

Like mode 1, mode 9 shows coordinated changes affecting craniofacial and cervical structures. Of particular interest here, however, are the coordinated changes affecting the VPO, the base of the tongue and dimensions of the hypopharyngeal airway. The existence of synergistic relationships between muscles controlling the size and shape of the VPO and hypopharynx receive strong support from recent anatomical studies demonstrating the presence of functional relationships between superior constrictor and airway dilator muscles (e.g. genioglossus). (25,49) Furthermore, the observation that changing hypopharyngeal dimensions reflect only part of more extensive coordinated pitch-related changes is especially interesting, particularly given the long-standing search for mechanisms underlying the production of the “singing formant” where the presence of greater spectral energy around 3000 Hz allows the singer’s voice to be heard over the sound of an orchestra. (7) It is suggested that conditions for production of the “singing formant” are met when the ratio between the diameters of the laryngeal tube and hypopharynx is 6:1. Positional changes of the tongue and lowering of the larynx are known to affect hypopharyngeal dimensions (7) but, as yet, a coherent explanation of the mechanisms underlying these changes is lacking. (27,50)

Our demonstration that changes of hypopharyngeal dimensions occur as part of a more extensive coordinated response illustrates the significant potential of active shape modelling to uncover mechanisms underlying the production of the ‘singing formant’, and explanations for vocal phenomena such the rise and fall of the larynx with changes of pitch. (8,9) Together, modes 1 and 9 demonstrate that widespread, coordinated pitch-related adjustments occur throughout the head and neck during pitch production, even in the absence of articulator input (lip and jaw). Like the results of geometric analysis reported previously (11,12), these findings challenge traditional theories of pitch production that rely on source-tract independence. Instead, they align more comfortably with older theories suggesting that the voice source and filter (vocal tract and supralaryngeal structures) are mutually interdependent and that the vocal ‘instrument’ should be considered as a whole. (51-55) This view receives support from converging evidence pointing to the importance of the upper cervical region in allowing the ‘organism’ to function as a finely coordinated whole. (56-58)

The importance of this method lies in its power to: 1) explain previous reports of synergy between vocal structures, illustrated by recent observations that pharyngeal constriction “almost always” occurs in parallel with other phenomena such as a “change in larynx height and a tendency to velar lowering.” (59); 2) account for a lack of synergy in situations where it was expected, such as the lack of correlation between hyoid and jaw movements in speech (60); and 3) to demonstrate that anatomical and functional variations involving individual vocal structures need to be considered in a wider context if important information is not to be missed because they reflect only part of the underlying coordinated whole. We suggest that the presence of an underlying, independently controlled pitch-adjusting system could explain the above observations.

The existence of a pitch-adjusting system that is integrated with the articulatory system but under independent control could offer a new and intriguing perspective from which to consider mechanisms underlying a wide range of speech and tonal phenomena: for example, how tonal differences affect supralaryngeal articulation (61), and the nature of speech production goals. (1) Recent demonstrations of the existence of separate pathways for the control of innate and learned vocalization patterns are consistent with this view. (28)

Mode 2

The results for mode 2, accounting for a fifth of total variance, were not significant, i.e., for this mode there were no significant differences in this score between each of the 6 conditions. However, lack of significance does not mean that this mode is not important as a source of valuable information. Evidence from orthodontic literature indicating the presence of coordinated patterns of growth affecting head and neck development suggest that this mode reflects the wide variation of individual head-neck shapes within this group of 10 subjects. Solow and Tallgren (1976), for example, reported an association between upper craniocervical angles and craniofacial dimensions (62) and, more recently, we reported correlations between the lower craniocervical angle and craniofacial dimensions. (11) The findings of this mode highlight the potential of this method to significantly advance knowledge and understanding of underlying coordinated patterns of head-neck development in health and disease. Consequently, these findings are also likely to be of interest to other disciplines interested in factors affecting the size and shape of the airway; orthodontics, maxillofacial surgery and sleep apnoea research, for example.

Overall, the findings of modes 1, 2 and 9 demonstrate the importance of investigating vocal structures and the vocal tract within the context of their wider structural relationships if important findings are not to be missed, thereby supporting Oudeyer's assertion (2005) that a wider focus, necessary to appreciate interactions of many components, could potentially "uncover major phenomena of speech and language." (63)

Limitations

This study has a number of limitations. It is based on a small sample. Increasing the number of participants would improve the capacity of this method to model variation within the chosen population. Vocal tract morphology, for example, is known to differ between men and women (e.g. vocal tract length is greater in men (4), and hyoid position is higher and more posterior in women (64)). With only 5 men and 5 women, numbers here are too small to draw meaningful conclusions about male/female differences; however, with a larger sample size, and its ability to uncover hidden shape patterns, active shape modelling offers a promising new approach in the search to understand factors underlying sex-related differences of vocal morphology. Manual annotation of the MR images is time consuming. As the number of images incorporated into the model increases, the model's ability to find and map on to the defined shape in a new image improves. However, even with 60 images incorporated into the model it is still necessary to precisely match each point of the new image to the flexible template. Without such fine adjustment, the likelihood of the model being able to register subtle synergistic activity involving key vocal tract regions (e.g., VPO and the width of the laryngeal tube opening) would be significantly diminished. The values assigned to the modes are not directly comparable with existing conventional geometric measurements. Nevertheless, by identifying how structures move in relation to each other, findings derived from this method complement and extend information obtained by using traditional measurements. Mode scores refer to variations about the mean for this particular set of images. This means that results obtained from this model cannot be directly compared with those obtained from a model based on a different set of images. Only one observer (NAM) annotated the images therefore more work needs to be done to establish the reliability of these findings. However, reports of low intra-investigator variability compared with inter-investigator variability suggest point placement by trained observers is reliable and, unsurprisingly, that preference should be given to such intra-investigator contributions when comparing a series of MR images. (65) Finally, the ASM was built using midsagittal MR images. Unlike coronal images, a midsagittal view does not show changes affecting the lateral wall of the vocal tract/airway which are known to be active in voice production. (8)

Implications and future work

Our findings demonstrate the significant potential of active shape modelling to advance knowledge and understanding of factors underlying anatomical and functional variations affecting the cervical spine, airway, and craniofacial and vocal structures, both during development and as a result of disease. We suggest that use of this method could lead to important new insights into causal mechanisms underlying such variations. In turn, this could assist our ability to quantify and interpret changes associated with voice production and, therefore, find answers for fundamental questions in voice science. Identification of coordinated mechanisms underlying vocal behavior could pave the way for more effective treatments and therapies for those with communication difficulties and, by looking beyond vocal tract geometry, the development of perceptually more accurate biomechanical models for voice synthesis.

Work is underway to further explore the use of this model. In previously published work, results obtained for the professional singer were opposite to those obtained from the rest of the group. As trained singers are encouraged to adopt a low larynx, it is possible that the switch from low- to high-note humming was associated with different modes of variation. To investigate this possibility further, we are repeating this study with professional singers.

As this model combines shape information from a number of individuals, we cannot draw conclusions about underlying causal mechanisms in any one individual. However, findings from ASMs could inform the choice of muscles targeted in electromyography experiments which, in turn, could enlighten our understanding of muscles underlying pitch-related phenomena such as the rise and fall of the larynx, and mechanisms that underlie the control of the VPO and hypopharynx. Improved knowledge of mechanisms underlying such coordinated activity has important implications for our understanding and, therefore, teaching of vocal techniques in professional singers. Recognition of the importance of the coordinating role of lower cranial nerves and upper cervical nerves in pitch-related activity, together with knowledge of structural attachments (e.g. of omohyoid to scapula) could also, we suggest, lead to a better understanding of mechanisms responsible for the close association of pitch with posture (66), gesture (67), and expression. (68)

Conclusion

Our results highlight active shape modelling's significant potential as an important method for identifying and distinguishing between structural and functional changes affecting the cervical spine, vocal structures and the airway. Use of this method can also show how different vocal structures move together as the overall shape varies according to development or function. By presenting results in a dynamic visual format, ASM not only complements findings of a previous study where more conventional measurements were used, it also extends them by demonstrating an unexpected and surprising association between pitch-related voice production and changes involving the cervical spine. Our results highlight the potential of active shape modelling to significantly advance knowledge and understanding of factors underlying structural and functional variations in health and disease. In turn, this could lead to better treatments and therapies for those with voice difficulties, more effective strategies for improving vocal technique in professional singers, and new insights and testable hypotheses for a wide range of vocal phenomena.

Acknowledgements

We are grateful to Dr Scott Semple and MRI radiographers for their help in acquiring images for earlier studies, volunteers for freely donating their time, Professor Alison D. Murray for reading an earlier draft of the paper, and the University of Aberdeen for sponsoring this study. We also extend our thanks to the anonymous reviewers for their insightful critique and suggestions.

References

(1) Munhall KG. Functional imaging during speech production. Acta Psychol 2001;107(95):117.

(2) Masaki S, Nota Y, Takano S, Takemoto H, Kitamura T, Honda K. Integrated magnetic resonance imaging methods for speech science and technology. J Acoust Soc Am 2008;123(5):3734.

(3) Narayanan S, Nayak K, Lee S, Sethy A, Byrd D. An approach to real-time magnetic resonance imaging for speech production. Journal of the Acoustical Society of America 2004;115(4):1771-1776.

(4) Fitch WT, Giedd J. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. Journal of the Acoustical Society of America 1999;106(3 I):1511-1522.

(5) Kitamura T, Honda K, Takemoto H. Individual variation of the hypopharyngeal cavities and its acoustic effects. Acoustical Science and Technology 2005;26(1):16-26.

(6) Moore RK. Spoken language processing: Piecing together the puzzle. Speech Commun 2007;49(5):418-435.

(7) Sundberg J. Articulatory interpretation of the “singing formant”. J Acoust Soc Am 1974;55(4):838-844.

(8) Yanagisawa E, Estill J, Mambrino L, Talkin D. Supraglottic contributions to pitch raising. Videoendoscopic study with spectroanalysis. Ann Otol Rhinol Laryngol 1991 Jan;100(1):19-30.

(9) Andrade PA. Analysis of male singers laryngeal vertical displacement during the first passaggio and its implications on the vocal folds vibratory pattern. J Voice 2012 Sep;26(5):665.e19-665.e24.

(10) Traser L, Burdumy M, Richter B, Vicari M, Echternach M. The effect of supine and upright position on vocal tract configurations during singing—a comparative study in professional tenors. Journal of Voice 2013;27(2):141-148.

(11) Miller NA, Gregory JS, Semple SIK, Aspden RM, Stollery PJ, Gilbert FJ. Relationships between vocal structures, the airway, and craniocervical posture investigated using magnetic resonance imaging. Journal of Voice 2012;26(1):102-109.

(12) Miller NA, Gregory JS, Semple SIK, Aspden RM, Stollery PJ, Gilbert FJ. The effects of humming and pitch on craniofacial and craniocervical morphology measured using MRI. Journal of Voice 2012;26(1):90-101.

(13) Anegawa E, Tsuyama H, Kusukawa J. Lateral cephalometric analysis of the pharyngeal airway space affected by head posture. International Journal of Oral & Maxillofacial Surgery 2008;37(9):805-809.

(14) Cootes T, Baldock ER, Graham J. An introduction to active shape models. Image Processing and Analysis 2000:223-248.

(15) Cootes TF, Hill A, Taylor CJ, Haslam J. Use of active shape models for locating structures in medical images. Image Vision Comput 1994;12(6):355-365.

(16) Cootes TF, Taylor CJ, Cooper DH, Graham J. Active shape models-their training and application. Comput Vision Image Understanding 1995;61(1):38-59.

(17) Cootes TF, Taylor CJ. Statistical models of appearance for medical image analysis and computer vision. Proc. SPIE 4322, Medical Imaging 2001:Image Processing, 236 (July 3) 2001.

(18) Cootes TF, Cooper DH, Taylor CJ. A trainable method of parametric shape description. Proceedings of the British Machine Vision Conference 1991.

(19) Barr RJ, Gregory JS, Reid DM, Aspden RM, Yoshida K, Hosie G, et al. Predicting OA progression to total hip replacement: can we do better than risk factors alone using active shape modelling as an imaging biomarker? Rheumatology 2012;51(3):562-570.

(20) Meakin JR, Gregory JS, Aspden RM, Smith FW, Gilbert FJ. The intrinsic shape of the human lumbar spine in the supine, standing and sitting postures: characterization using an active shape model. J Anat 2009;215(2):206-211.

(21) Vasconcelos MJM, Ventura SMR, Tavares JMRS, Freitas DRS. Analysis of tongue shape and motion in speech production using statistical modeling. 2nd South-East European Conference on Computational Mechanics 2009.

(22) Vasconcelos MJM, Ventura SMR, Freitas DRS, Tavares JMRS. Towards the automatic study of the vocal tract from magnetic resonance images. Journal of Voice 2011;25(6):732-742.

(23) Ventura SMR, Vasconcelos MJM, Freitas DRS, Ramos IMAP, Tavares JMRS. Speaker-specific articulatory assessment and measurements during Portuguese speech production based on magnetic resonance Images. In: Warfelt LM, editor. Language acquisition: Gazelle; 2012.

(24) Vasconcelos MJM, Ventura SMR, Freitas DRS, Tavares JMRS. Inter-speaker speech variability assessment using statistical deformable models from 3.0 Tesla magnetic resonance images. Proc Inst Mech Eng Part H J Eng Med 2012;226(3):185-196.

(25) Saigusa H, Yamashita K, Tanuma K, Saigusa M, Niimi S. Morphological studies for retrusive movement of the human adult tongue. Clin Anat 2004 Mar;17(2):93-98.

(26) Standring S, Ellis H, Healy JC, et al editors. Gray's anatomy: the anatomical basis of clinical practice. 39th ed. Edinburgh, UK: Churchill Livingstone; 2005.

(27) Imagawa H, Sakakibara K, Tayama N, Niimi S. The effect of the hypopharyngeal and supra-glottic shapes on the singing voice. Proceedings of the Stockholm Music Acoustics Conference 2003 2003;3:471-474.

(28) Simonyan K, Horwitz B. Laryngeal motor cortex and control of speech in humans. Neuroscientist 2011;17(2):197-208.

(29) Zarate JM. The neural control of singing. Frontiers in human neuroscience 2013;7.

(30) Fowler LP, Morris RJ. Comparison of nasalance between trained singers and non-singers. Electronic Theses, Treatises and Dissertations. paper 4411 2004.

(31) Honda, H. Physiological factors causing tonal characteristics of speech: from global to local prosody. Speech Prosody 2004, International Conference: ISCA; 2004.

(32) Vilkman E, Sonninen A, Hurme P, Körkkö P. External laryngeal frame function in voice production revisited: a review. Journal of Voice 1996;10(1):78-92.

(33) Miller AJ. Oral and pharyngeal reflexes in the mammalian nervous system: their diverse range in complexity and the pivotal role of the tongue. Crit Rev Oral Biol Med 2002;13(5):409-425.

(34) Mitchinson AGH, Yoffey JM. Changes in the vocal folds in humming low and high notes. A radiographic study. J Anat 1948;82(Pt 1-2):88.

(35) Perera L, Kern M, Hofmann C, Tatro L, Chai K, Kuribayashi S, et al. Manometric evidence for a phonation-induced UES contractile reflex. American Journal of Physiology-Gastrointestinal and Liver Physiology 2008;294(4):G885-G891.

(36) Rossum MA, Krom G, Nooteboom SG, Quene H. " Pitch" accent in alaryngeal speech. Journal of Speech, Language and Hearing Research 2002;45(6):1106-1118.

(37) Honda K, Hirai H, Masaki S, Shimada Y. Role of vertical larynx movement and cervical lordosis in F0 control. Lang Speech 1999;42(4):401.

(38) Fant G. Acoustic theory of speech production. : Mouton De Gruyter; 1970.

(39) Scotto Di Carlo N. Cervical column abnormalities in professional opera singers. Folia Phoniatr Logop 1998;50:212-218.

(40) Johnson G, Skinner M. The demands of professional opera singing on cranio-cervical posture. Eur Spine J 2009 Apr;18(4):562-569.

(41) Kitamura T, Takemoto H, Honda K, Shimada Y, Fujimoto I, Syakudo Y, et al. Difference in vocal tract shape between upright and supine postures: observations by an open-type MRI scanner. Acoustical Science and Technology 2005;26(5):465-468.

(42) Mayoux-Benhamou MA, Revel M, Vallee C, Roudier R, Barbet JP, Bargy F. Longus colli has a postural function on cervical curvature. Surgical and Radiologic Anatomy 1994;16(4):367-371.

(43) Yamawaki Y. Forward movement of posterior pharyngeal wall on phonation. Am J Otolaryngol 2003 Nov-Dec;24(6):400-404.

(44) Pettersen V, Westgaard RH. The activity patterns of neck muscles in professional classical singing. Journal of Voice 2005;19(2):238-251.

(45) Chusid JG. Correlative neuroanatomy and functional neurology. 17th ed. Los Altos, CA: Lange Medical Publications; 1979.

(46) Roubeau B, Chevrie-Muller C, Saint Guily JL. Electromyographic activity of strap and cricothyroid muscles in pitch change. Acta Otolaryngol 1997;117(3):459-464.

(47) Van Houtte E, Van Lierde K, Claeys S. Pathophysiology and treatment of muscle tension dysphonia: a review of the current knowledge. Journal of Voice 2011;25(2):202-207.

(48) Vitti M, Fujiwara M, Basmanjian JM, Iida M. The integrated roles of longus colli and sternocleidomastoid muscles: an electromyographic study. Anat Rec 1973 Dec;177(4):471-484.

(49) Kokawa T, Saigusa H, Aino I, Matsuoka C, Nakamura T, Tanuma K, et al. Physiological studies of retrusive movements of the human tongue. J Voice 2006 Sep;20(3):414-422.

(50) Leino T, Laukkanen A, Radolf V. Formation of the actor's/speaker's formant: a study applying spectrum analysis and computer modeling. Journal of Voice 2011;25(2):150-158.

(51) Bell C. Of the Organs of the Human Voice. Philosophical Transactions of the Royal Society of London 1832;122:299-320.

(52) Bishop J. On the Physiology of the Human Voice. Philosophical Transactions of the Royal Society of London 1846;136:551-571.

(53) Browne L. On Medical Science in Relation to the Voice as a Musical Instrument. Proceedings of the Musical Association 1875;2:94-110.

(54) Steed AO. On Beauty of Touch and Tone: An Inquiry into the Physiological and Mechanical Principles Involved in Their Cultivation. Part I. The Voice. Proceedings of the Musical Association 1879;6:31-58.

(55) Stanley D. The science of voice. Journal of the Franklin Institute 1931;211(4):432-436.

(56) Eriksson PO. Co-ordinated mandibular and head-neck movements during rhythmic jaw activities in man. J Dent Res 2000;79(6):1378-1384.

(57) Brown H, Hidden G, Ledroux M, Poitevan L. Anatomy and blood supply of the lower four cranial and cervical nerves: relevance to surgical neck dissection. Exp Biol Med 2000;223(4):352-361.

(58) Morch CD, Hu JW, Arendt-Nielsen L, Sessle BJ. Convergence of cutaneous, musculoskeletal, dural and visceral afferents onto nociceptive neurons in the first cervical dorsal horn. Eur J Neurosci 2007;26(1):142-154.

(59) Hanayama EM, Camargo ZA, Tsuji DH, Rebelo Pinho SM. Metallic voice: physiological and acoustic features. Journal of Voice 2009;23(1):62-70.

(60) Hiiemae KM, Palmer JB, Medicis SW, Hegener J, Scott Jackson B, Lieberman DE. Hyoid and tongue surface movements in speaking and eating. Arch Oral Biol 2002;47(1):11-27.

(61) Hu F. Tonal effect on vowel articulation in a tone language. International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages 2004.

(62) Solow B, Tallgren A. Head posture and craniofacial morphology. Am J Phys Anthropol 1976 May;44(3):417-435.

(63) Oudeyer P. The self-organization of speech sounds. J Theor Biol 2005;233(3):435-449.

(64) Sahin Saglam AM, Uydas NE. Relationship between head posture and hyoid position in adult females and males. Journal of cranio-maxillo-facial surgery 2006;34(2):85-92.

(65) Stuck BA, Maurer JT. Airway evaluation in obstructive sleep apnea. Sleep Medicine Reviews 2008;12(6):411-436.

(66) Kuratate T, Munhall KG, Rubin PE, Vatikiotis-Bateson E, Yehia H. Audio-visual synthesis of talking faces from speech production correlates. Sixth European Conference on Speech Communication and Technology 1999:1279-1282.

(67) McClave E. Pitch and manual gestures. J Psycholinguist Res 1998;27(1):69-89.

(68) Huron D, Shanahan D. Eyebrow movements and vocal pitch height: Evidence consistent with an ethological signal. J Acoust Soc Am 2013;133:2947.

Revised ASM paper

Figures, Table and Legends

Nicola Miller 19th November 2013

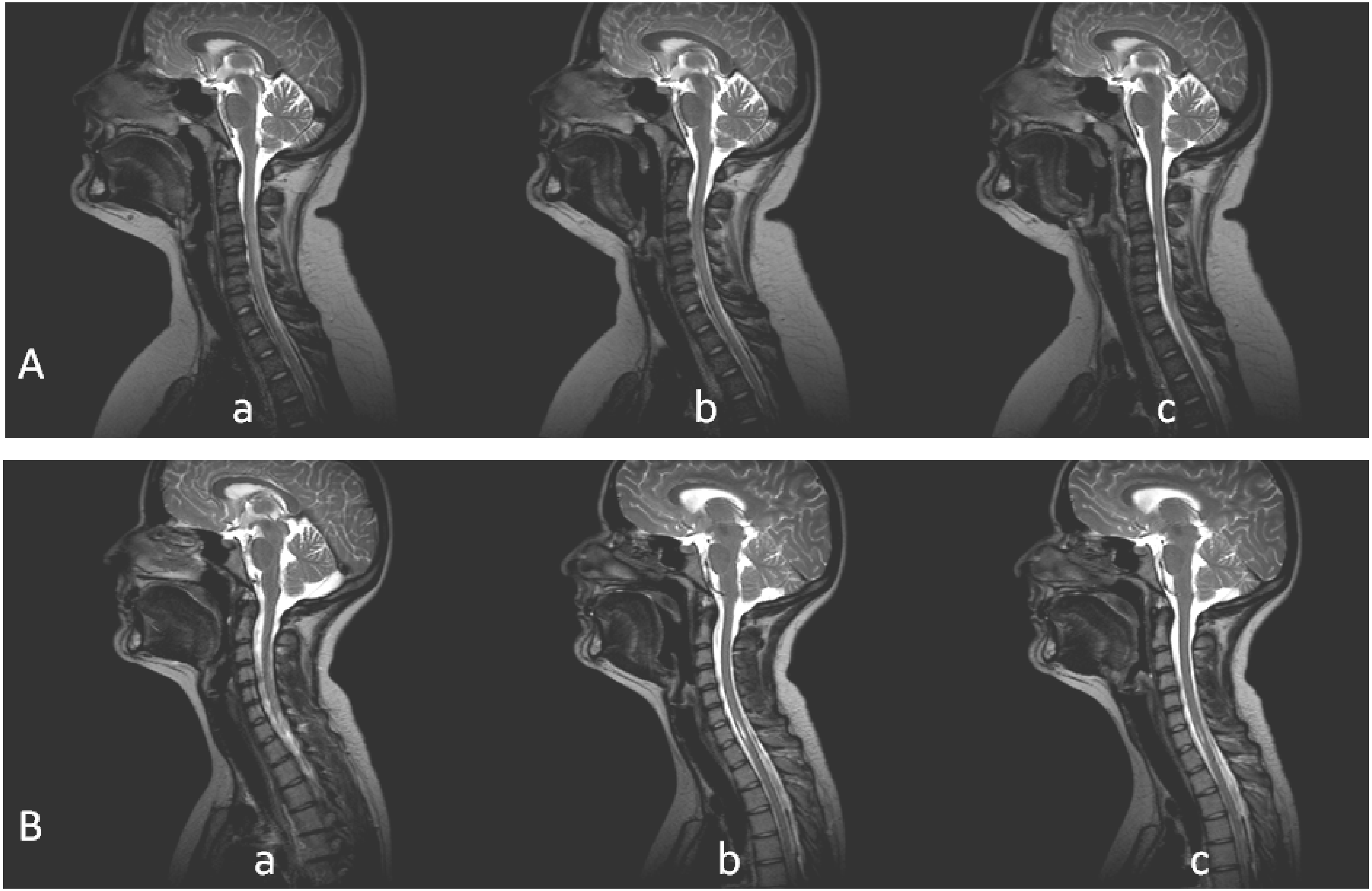
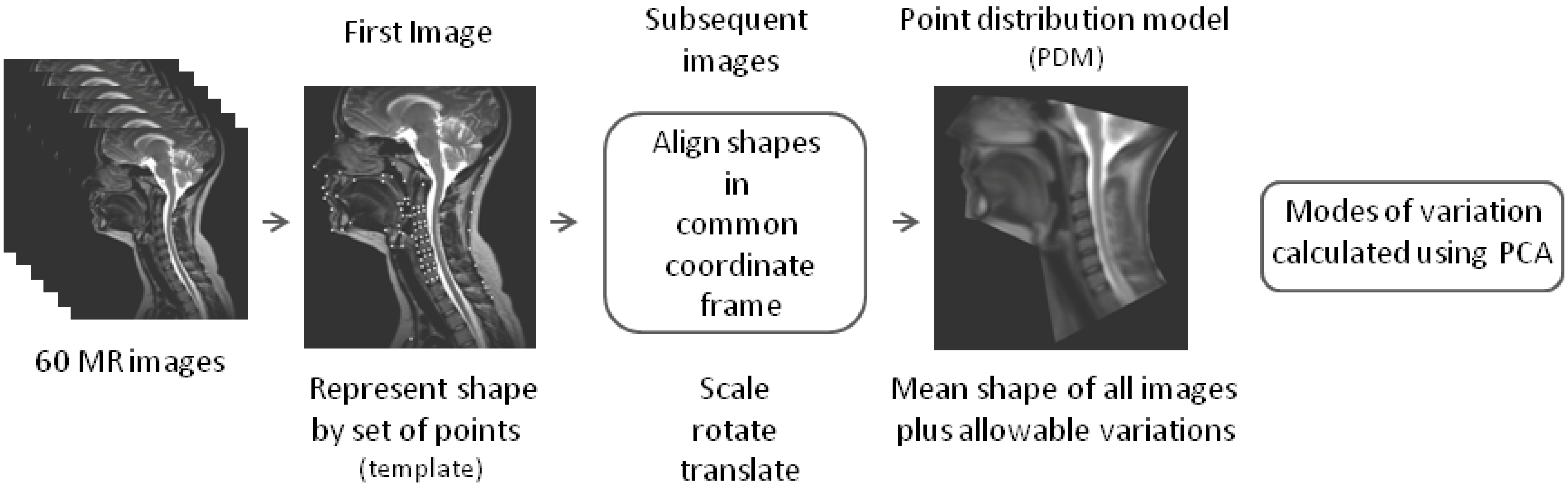


Figure 1

Midsagittal MR images from 2 subjects, A and B, acquired a) while at rest b) while humming a low note and c) while humming a high note. Note differences between subjects at rest (e.g. of cervical alignment), and differences within and between subjects humming low and high notes (e.g. of cervical alignment, larynx height, tongue shape and soft palate).

Figure 2



60 MR images

First Image

Represent shape by set of points (template)

Subsequent images

Align shapes in common coordinate frame

Scale rotate translate

Point distribution model (PDM)

Mean shape of all images plus allowable variations

Modes of variation calculated using PCA

Summary of steps taken to build an active shape model

Figure 3

Figure 1. Typical midsagittal MR image showing positions of the 80 landmark points used to define the template of the active shape model representing vocal structures, the airway and the cervical spine (C2-C7)

Table 1

Mode of variance	**Retained variance %**	**Cumulative variance %**
1	30.64	31
2	20.39	51
3	12.65	64
4	8.13	72
5	6.24	78
6	4.33	82
7	3.68	86
8	2.87	89
9	1.82	91
10	1.34	92
11	0.98	93
12	0.86	94
13	0.75	95
14	0.71	95
15	0.59	96
16	0.56	97
17	0.40	97
18	0.30	97
19	0.27	98
20	0.26	98

Table of modes of variation and percentage variance

Figure 4

Mean shape illustrating key anatomical features

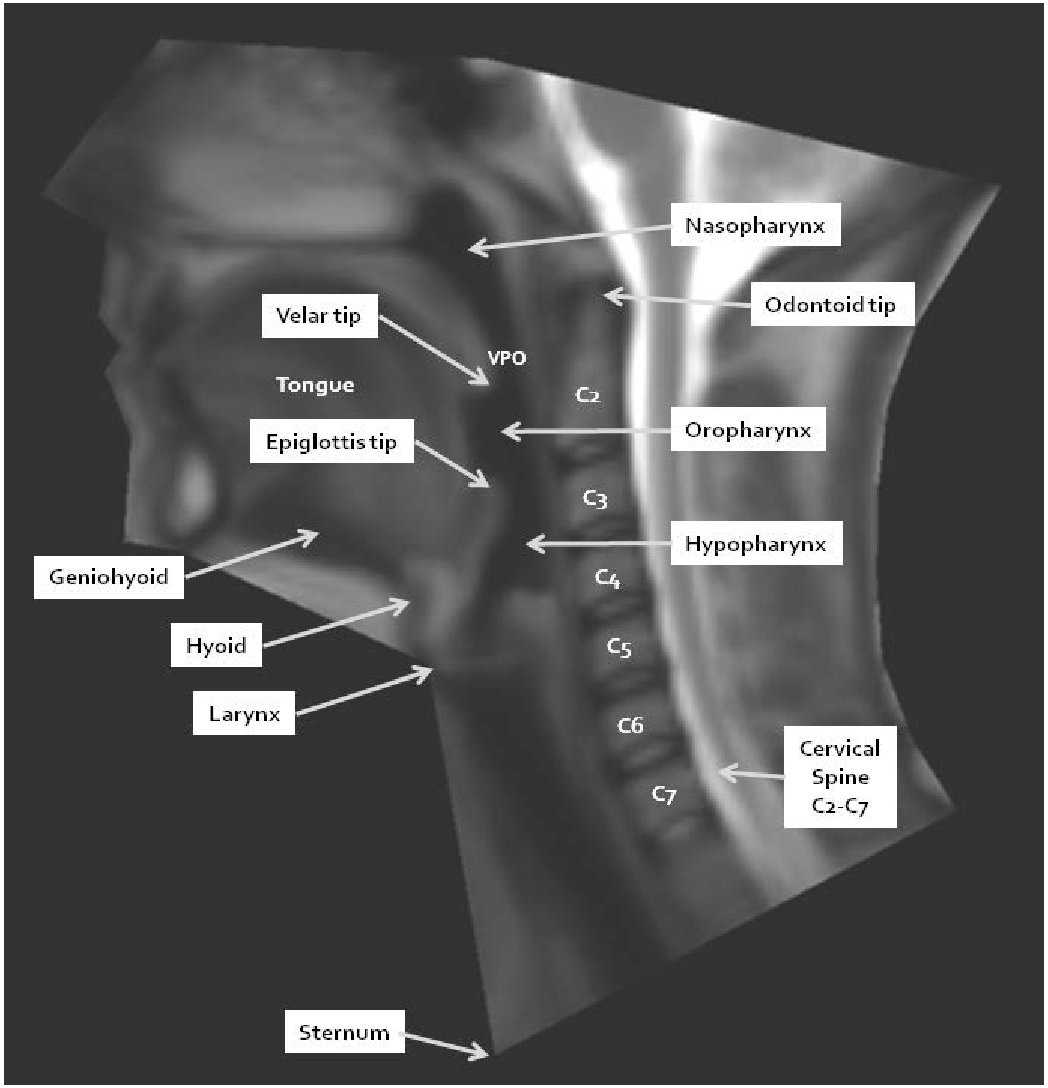


Figure 5. Shapes described by varying Mode 1 (5a) and Mode 9 (5b) by +2 to -2 standard deviations about the mean shape of all 60 images. Click images to see shape animations

Figure 5a Mode 1

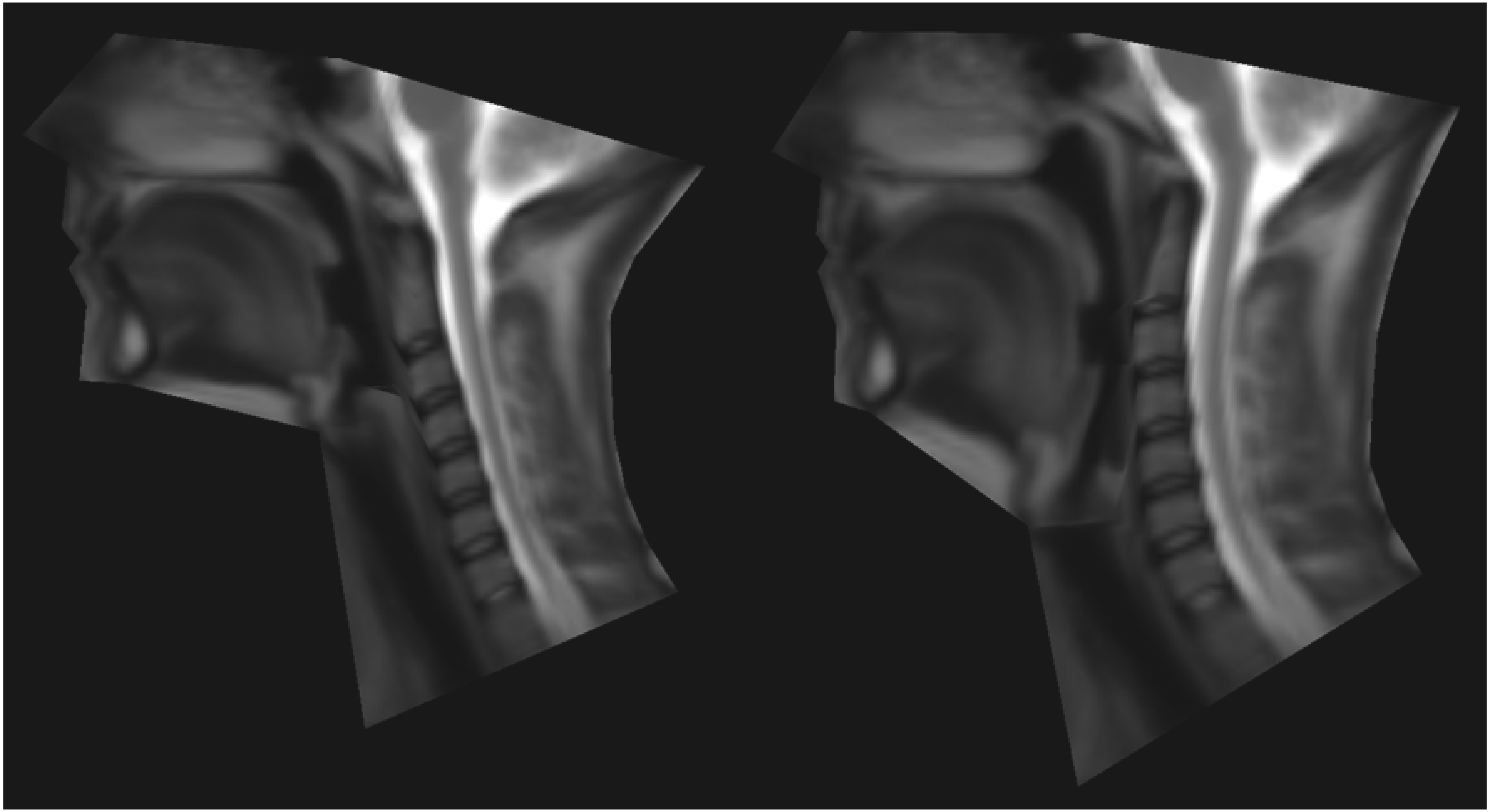


Figure 5b Mode 9

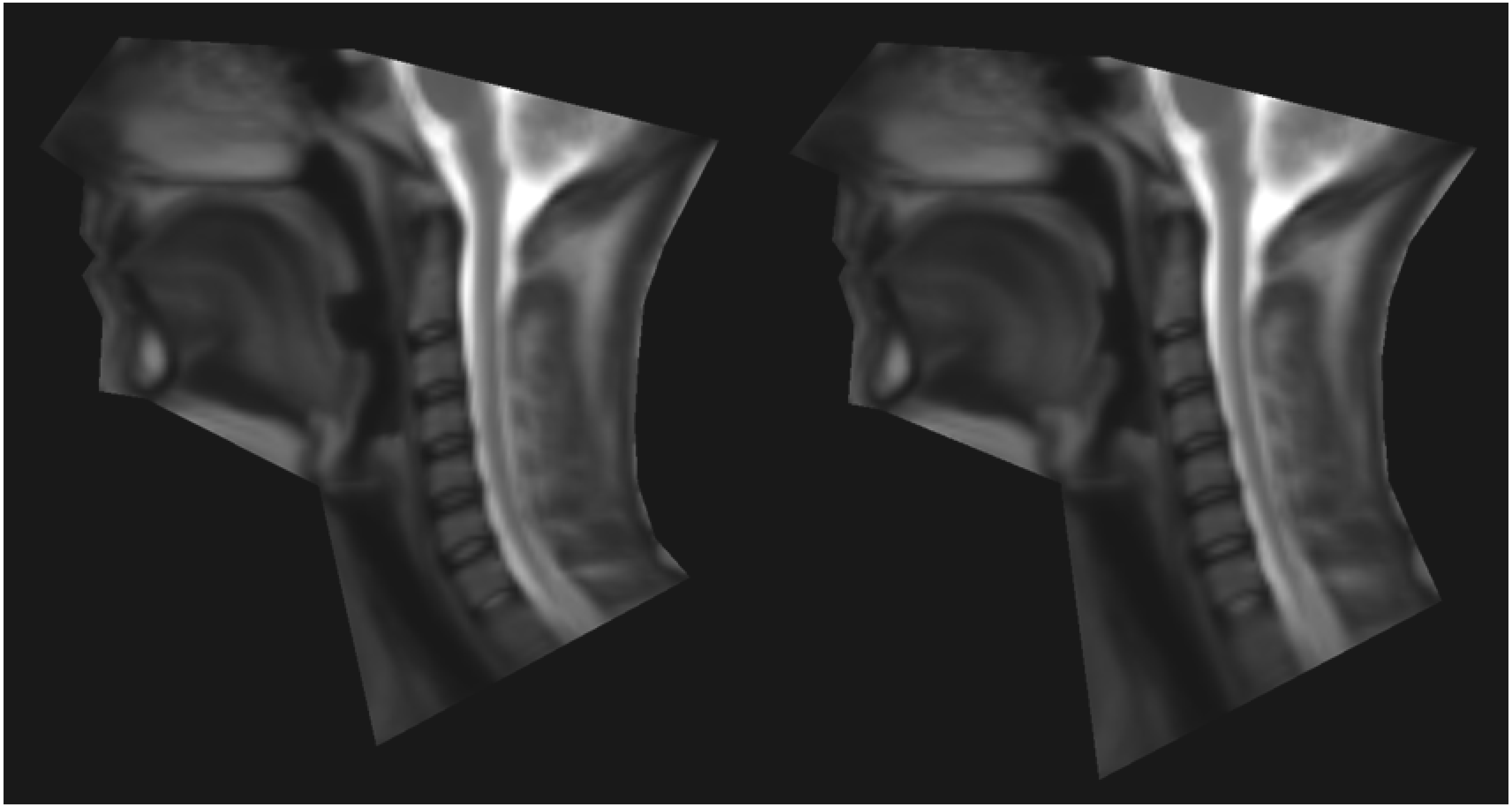


Figure 6. Shapes described by varying Mode 2 by +2 to –2 standard deviations about the mean shape of all 60 images. Click image to see shape animation.

.

Figure 6 Mode 2