Descent of the hyoid in chimpanzees: evolution of face flattening and speech

Takeshi Nishimura a,*, Akichika Mikami b, Juri Suzuki b, Tetsuro Matsuzawa b

a Laboratory of Physical Anthropology, Department of Zoology, Graduate School of Science, Kyoto University,
Kitashirakawa Oiwake-cho, Sakyo, Kyoto 606-8502, Japan
b Primate Research Institute, Kyoto University, Inuyama, Aichi 484-8506, Japan

Received 1 September 2005; accepted 27 March 2006

Abstract

The human supralaryngeal vocal tract develops to form a unique two-tube configuration with equally long horizontal and vertical cavities. This anatomy contributes greatly to the morphological foundations of human speech. It is believed to depend on the reduced growth of the palate and on the developmental descent of the larynx relative to the palate. Anatomically, the descent of the larynx is accomplished through both the descent of the laryngeal skeleton relative to the hyoid and the descent of the hyoid relative to the palate. We have studied the development of three living chimpanzees using magnetic resonance imaging. Our previous study showed that, as in humans, chimpanzees show rapid laryngeal descent, with changes in the relative proportion of the vocal tract during early infancy. However, this is not accompanied by the descent of the hyoid relative to the palate, although it is achieved with the descent of the laryngeal skeleton relative to the hyoid. Here, we show that subsequently the chimpanzee hyoid also descends to maintain the rapid descent of the larynx, similarly to humans. We argue that the descent of the larynx probably evolved in a common ancestor of extant hominoids, originally to confer an advantage via a function unrelated to speech. Thus, the descent of the larynx per se is not unique to humans, and facial flattening was probably the major factor that paved the way for speech in the human lineage.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Descent of the larynx; Vocal tract; Pan troglodytes; MRI; Longitudinal study

Introduction

The evolution of human speech has attracted much interest to better understand the evolution of language. Human speech has the distinct feature that humans can regularly utter several phonemes—including vowels and consonants—sequentially and rapidly in a short, single exhalation. It must be noted that speech per se is not the same as language and does not necessarily reflect the high intelligence of humans. However, this sophisticated feature of speech allows humans to turn much information that is encoded by language in the brain into sounds and to communicate it to others rapidly and efficiently (P. Lieberman, 1984). Therefore, even if language and speech arose independently in the human lineage, an understanding of the evolution of speech would shed light on the evolution of language with which we are endowed today.

Theories of speech physiology and acoustics demonstrate that humans have a unique anatomy of the supralaryngeal vocal tract (SVT) that underlies its sophisticated manipulation in the production of speech. Humans and nonhuman mammals basically make use of the same machinery for speech and vocalization: the lungs generate sound power, the vocal folds in the larynx comprise the sound sources, and the SVT resonates the sources to generate voiced sounds with some bands of the formant frequencies, e.g., vowels in speech (Fant, 1960; P. Lieberman and Blumstein, 1988; Titze, 1994; Stevens, 1998; Fitch and Hauser, 2003; Riede et al., 2005). The
distribution pattern of the formants defines the different kinds of vowels that form the platform for vocal communication. These are determined by the resonance properties of the SVT and these properties in turn are dictated by the volumetric topology of the tract, which can be estimated from a function of the sequential cross-sectional area along the tract. The SVT in most mammals, including humans, is principally composed of two cavities: the horizontal oral and vertical pharyngeal cavities designated here the SVTH and SVTV, respectively. In nonhuman mammals, the SVTV is much shorter than the SVTH (Negus, 1949; P. Lieberman, 1984; Laitman and Reidenberg, 1993; Dyce et al., 1996). Their epiglottis, which is attached to the thyroid cartilage of the laryngeal skeleton, maintains contact with the velum and prevents the SVTV from facing the movable tongue. The tongue is also long in the horizontal direction, fitting this configuration. Although this anatomy allows the oral cavity to function as a single resonator, it prevents the pharyngeal cavity from contributing much in that capacity (P.H. Lieberman et al., 1969; P. Lieberman, 1984; Fitch, 2000a; Fitch and Hauser, 2003). Thus, nonhuman mammals have physical constraints to any rapid sequential modification of the cross-sectional areas of the SVT. In contrast, adult humans have an equally long SVTH and SVTV. Their epiglottis is separated from the velum, and this produces a long oropharyngeal region facing the dorsal surface of the tongue, rostral to the laryngopharyngeal region that faces the epiglottis (P. Lieberman, 1984; Crelin, 1987; Zemlin, 1988; Titze, 1994). The vertical dimension of the tongue is almost equal to the horizontal dimension to fit this configuration, and the internal musculature of the tongue makes the surface highly mobile (Takemoto, 2001). In anatomical terms, these features allow the shapes of the SVTH and SVTV to be sequentially and rapidly modified by tongue movements, semi-independently of each other (P.H. Lieberman et al., 1969; P. Lieberman, 1984; Fitch, 2000a; Fitch and Hauser, 2003).

Thus, humans are capable of extensive modification of the resonant properties of the SVT, which in turn modify the laryngeal sounds, forming the complex sequential phonemes of speech in a single, short exhalation.

In newborn humans, the larynx is positioned close to the palate so there is little vertical pharyngeal space (Negus, 1949; P. Lieberman, 1984; Crelin, 1987; D.E. Lieberman et al., 2001). However, the major descent of the larynx causes the epiglottis to descend relative to the velum (Negus, 1949; Roche and Barkla, 1965; Sasaki et al., 1977; Crelin, 1987; Westhorpe, 1987; Fitch and Giedd, 1999; Vorperian et al., 1999, 2005). This establishes a long oropharyngeal space, rostral to the laryngopharyngeal region in the SVTV. Thus, the SVTV lengthens rapidly compared with the SVTH, and can function as a resonance tube of equivalent volume (P. Lieberman, 1984; Crelin, 1987; Titze, 1994; Fitch, 2000a).

The laryngeal skeleton is suspended from the hyoid apparatus, and the hyoid is in turn suspended from the mandible and cranial base by muscles and ligaments (Zemlin, 1988). Anatomically, two processes accomplish the descent of the larynx: the descent of the laryngeal skeleton relative to the hyoid, and the descent of the hyoid relative to the palate. This process has been evaluated with detailed measurements on humans from X-ray photographs (D.E. Lieberman and McCarthy, 1999; D.E. Lieberman et al., 2001) and magnetic resonance imaging (MRI; Fitch and Giedd, 1999; Vorperian et al., 1999, 2005).

We have studied the development of three living chimpanzees using MRI. Our previous study compared the growth of the SVT in chimpanzees and humans during the first two years of life, and showed that—as in humans—chimpanzees show rapid laryngeal descent, with changes in the relative proportion of the SVTH and SVTV (Nishimura et al., 2003). Another MRI study (Nishimura, 2005), using a cross-sectional ontogenetic series of embalmed specimens, confirmed this developmental change during early infancy. However, that study also showed that the larynx is lowered only slightly and the horizontal oral cavity grows largely during the juvenile period, causing the proportion of the SVT in chimpanzees to differ from that in humans. Unfortunately, embalming artifacts precluded the study of developmental changes in the position of the hyoid and epiglottis relative to the palate. Therefore, the descent of the chimpanzee larynx is considered to depend primarily on the descent of the laryngeal skeleton relative to the hyoid, but not on the descent of the hyoid per se (Nishimura et al., 2003; Nishimura, 2005). Thus, the human 1:1 proportion of the SVTH to SVTV is still believed to depend on the greater proportional descent of the hyoid during the evolution of the human lineage (Negus, 1949; P. Lieberman, 1984; Crelin, 1987; Flügel and Rohren, 1991; Houghton, 1993; Laitman and Reidenberg, 1993; Nishimura, 2003, 2005; Nishimura et al., 2003). In fact, the term “descent” of the larynx or hyoid often implies the proportional changes of SVT toward such a configuration (here we use the terms “descent” and “descend” without any such implication). Here, we used MRI to evaluate the developmental changes in SVT anatomy in three living chimpanzees aged between two-and-a-half and five years, and we use the results to discuss the evolution of the descent of the larynx.

Methods

MRI procedures

We have studied the development of three living chimpanzees, named Ayumu (male), Cleo (female), and Pal (female), using MRI. They were born in 2000 and were reared by the biological mothers in the Primate Research Institute (PRI), Kyoto University, Japan (Matsuzawa, 2003). The care and use of the subjects conformed to the guidelines of the PRI (1986, 2002). The chimpanzees were scanned at scheduled intervals from four months to five years of age (Appendix 1). Here, we evaluate the measurements made from two-and-a-half to five years. The scans for Ayumu at four and six months of age were excluded from the analyses in our previous study (Nishimura et al., 2003) because the images were slightly obscured by motion artifacts.

The MRI examination protocol used here was essentially the same as described in a previous study (Nishimura et al., 2003) and was approved by the Ethics Panel of the PRI.
Sagittal tomographic images of the growing chimpanzee head and neck were taken with a General Electrics Signa Profile MRI scanner (0.2 Tesla) at the PRI, using the extremity or head receiving-coil. The subjects were anesthetized intramuscularly with 3.5 mg ketamine hydrochloride (Sankyo Co., Ltd., Tokyo, Japan) and 0.035 mg medetomidine hydrochloride (Meiji Seika Kaisha Ltd., Tokyo, Japan) per kilogram of body weight. After four years of age, they were sedated orally using 3.75 mg of droperidol (in 1.5 ml) before being anesthetized. They were placed supine with their heads fixed to the custom acrylic plastic pad within the coil with belts, so they were always placed in the same posture for scanning. The mid-sagittal plane of the head was adjusted to the laser lights of the MR scanner. This procedure inevitably bends the neck slightly at the level inferior to the vocal folds (Fig. 1). All imaging sequences were sagittal spin echo series with time-to-echo durations ranging from 17 to 32 ms, time-to-repeat durations of 600 ms, fields of view ranging from 18 to 28 cm with 2.7 mm or 3.0 mm slice thicknesses, and 0.8 mm or 0.5 mm gaps between slices using an acquisition matrix of 192 × 192 with two excitations. These parameters were chosen generally based on subject size (Appendix 1). The matrix of all MR images was 256 × 256 pixels, and image resolutions ranged from 0.70 × 0.70 to 1.09 × 1.09 mm/pixel (Appendix 1).

Measurements

MR images were transferred from the scanner to a personal computer, using Vox-Base Transmit software (J-Mac System, Sapporo, Japan). The images were converted from DICOM to TIFF format, using Intage software (KGT, Tokyo, Japan). These images were evaluated using Adobe PhotoShop CS software (Adobe Systems, San Jose, CA, USA) to record measurement points and standard planes. Measurement points and standard planes on the mid-sagittal plane included the following anatomical landmarks: anterior nasal spine (ANS); anterior tubercle of the atlas (ATA); endoprosthion (EPr); hyoid bone (HB); posterior nasal spine (PNS); posterior oropharyngeal wall (POW); palatal plane (PP); posterior pharyngeal wall line (PPW); and vocal fold (VF). These definitions (Table 1) are those defined for chimpanzees by Nishimura (2005) and are roughly equivalent to those used for radiographic studies of humans (D.E. Lieberman et al., 2001). The present study used two additional points: epiglottis (EG) and velum (VL). These definitions (given in Table 1) were those used for chimpanzees by Nishimura et al. (2003).

The coordinate values of the measurement points for linear dimensions were measured three times from MR images.
Table 1
Definitions of the measurement points and standard planes used

<table>
<thead>
<tr>
<th>Landmarks and planes</th>
<th>Abbr.</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior nasal spine</td>
<td>ANS</td>
<td>The most anterior inferior point of the piriform aperture of the nose, which is roughly equivalent to the most anterior root point of the nasal septum&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>Anterior tubercle of the atlas</td>
<td>ATA</td>
<td>The superior—inferior midpoint of the anterior tubercle of the atlas&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Epiglottis</td>
<td>EG</td>
<td>The most superior point of the epiglottis</td>
</tr>
<tr>
<td>Endoprophosphion</td>
<td>EPt</td>
<td>The most anterior inferior point of the lingual surface of the alveolar premaxilla</td>
</tr>
<tr>
<td>Posterior nasal spine</td>
<td>PNS</td>
<td>The most posterior point of the maxillary body at the level of the nasal floor at the junction of the hard and soft palates</td>
</tr>
<tr>
<td>Posterior margin of the oral cavity</td>
<td>POC</td>
<td>Plane from PNS to EPt—POW, parallel to the posterior pharyngeal wall (PPW)</td>
</tr>
<tr>
<td>Posterior oropharyngeal wall</td>
<td>POW</td>
<td>Point on the posterior pharyngeal wall opposite the anterior tubercle of the atlas along the plane from EPt to ATA</td>
</tr>
<tr>
<td>Palatal plane</td>
<td>PP</td>
<td>The line from the ANS to PNS</td>
</tr>
<tr>
<td>Posterior pharyngeal wall</td>
<td>PPW</td>
<td>The most posterior straight line of the pharyngeal wall from the level of the velum to the laryngeal orifice</td>
</tr>
<tr>
<td>Vocal folds</td>
<td>VF</td>
<td>The anterior—posterior midpoint of the vocal folds</td>
</tr>
<tr>
<td>Velum</td>
<td>VL</td>
<td>The most posterior inferior point of the palatine velum, excluding the uvula</td>
</tr>
</tbody>
</table>

All landmarks and planes were defined on the mid-sagittal plane of the MR images.

1 This definition is the same as that for chimpanzees used by Nishimura (2005) and Nishimura et al. (2003), but is different from that used by D.E. Lieberman et al. (2001) in the narrow sense because no true anterior nasal spine is found in chimpanzees.

2 This definition is the same as that used for chimpanzees by Nishimura (2005) and Nishimura et al. (2003), but is different from that used for the human ATA (D.E. Lieberman et al., 2001). The human ATA is defined as the most anterior and superior point of the anterior tubercle of the atlas that projects on the sagittal plane in radiographs. This point was identified only obscurely on mid-sagittal MR images of chimpanzee subjects, so the chimpanzee ATA was here redefined as being roughly equivalent to the human ATA.

The measurements were recorded for the chimpanzees at ages ranging from four months to five years (Nishimura et al., 2003; present study). Those for humans by D.E. Lieberman et al. (2001) were taken from one month to thirteen years and nine months of age. The developmental patterns were compared in terms of the developmental ages of both species after their chronological ages were adjusted to correspond to dental eruption stages, which reflect analogous growth phases in both species. In this study, three growth stages, “early infancy”, “late infancy”, and the “juvenile period”, were defined as the dental stage from birth to before the eruption of the deciduous dentition, the stage from the eruption of the deciduous dentition to before the eruption of the first molars, and the stage after the eruption of the first molars and before the eruption of the second molars, respectively. Based on studies of dental eruption ages in captive living chimpanzees (Conroy and Mahoney, 1991; Kuykendall et al., 1992) and European-derived humans (Hurme, 1949; Lysell et al., 1962; Smith and Garn, 1987; Smith, 1991), the chronological ages of the chimpanzees and humans were adjusted to correspond to dental stages. In chimpanzees, “early infancy” is defined here as the period from birth to one year of age, “late infancy” as the period from one to three years of age, and the “juvenile” stage as the period from three to seven years of age. In humans, “early infancy” is defined as the period from birth to two-and-a-half years of age, “late infancy” as the period from two-and-a-half to six years of age, and the “juvenile” stage as the period from six to twelve years of age (Fig. 2). Thus, the developmental patterns were examined here up to five years of age in chimpanzees and compared with those published for up to nine years of age in humans (Fig. 2).

Comparing growth phases between chimpanzees and humans

The measurements were recorded for the chimpanzees at ages ranging from four months to five years (Nishimura et al., 2003; present study). Those for humans by D.E. Lieberman et al. (2001) were taken from one month to thirteen years and nine months of age. The developmental patterns were compared in terms of the developmental ages of both species after their chronological ages were adjusted to correspond to dental eruption stages, which reflect analogous growth phases in both species. In this study, three growth stages, “early infancy”, “late infancy”, and the “juvenile period”, were defined as the dental stage from birth to before the eruption of the deciduous dentition, the stage from the eruption of the deciduous dentition to before the eruption of the first molars, and the stage after the eruption of the first molars and before the eruption of the second molars, respectively. Based on studies of dental eruption ages in captive living chimpanzees (Conroy and Mahoney, 1991; Kuykendall et al., 1992) and European-derived humans (Hurme, 1949; Lysell et al., 1962; Smith and Garn, 1987; Smith, 1991), the chronological ages of the chimpanzees and humans were adjusted to correspond to dental stages. In chimpanzees, “early infancy” is defined here as the period from birth to one year of age, “late infancy” as the period from one to three years of age, and the “juvenile” stage as the period from three to seven years of age. In humans, “early infancy” is defined as the period from birth to two-and-a-half years of age, “late infancy” as the period from two-and-a-half to six years of age, and the “juvenile” stage as the period from six to twelve years of age (Fig. 2). Thus, the developmental patterns were examined here up to five years of age in chimpanzees and compared with those published for up to nine years of age in humans (Fig. 2).

Results

The scan for Ayumu at 52 months of age showed that, at the time of scanning, the tongue was contracted anterosuperiorly, probably causing additional elevation of the hyoid and larynx (Fig. 3; Appendix 2). Although this indicates the dynamic anatomy of the chimpanzee vocal tract (discussed below), the scan was excluded from the analyses of the developmental changes in resting anatomy of the SVT.
In chimpanzees, the SVTV increases more than the SVTH during early infancy (Nishimura et al., 2003), but thereafter the SVTH shows greater growth relative to that of the SVTV (Fig. 1). This is reflected in the age-related changes in the ratio of the SVTH to SVTV lengths shown in Figure 4a and Appendix 2. Although the absolute values are greater in chimpanzees than in humans, the ratio decreases in chimpanzees—as in humans—during early infancy (Nishimura et al., 2003). After late infancy, it consistently decreases to reach the adult ratio of 1:1 by nine years of age in humans (D.E. Lieberman et al., 2001), whereas in chimpanzees the direction of change in the ratio reverses, and it increases to reach the ratio of 2.10 at five years of age. Figure 4b and Appendix 2 show...
the growth changes in \(SVT_V\) and \(SVT_H\) lengths measured by MRI in living chimpanzees. During early infancy, the growth patterns of these dimensions in chimpanzees are similar to those in humans, although the absolute values of the \(SVT_V\) are smaller in chimpanzees than in humans (Nishimura et al., 2003). Thereafter, the \(SVT_H\) in humans grows much more slowly (D.E. Lieberman et al., 2001), whereas in chimpanzees it continues to increase greatly and reaches 9.8 cm at five years. In contrast, even after late infancy, the growth pattern of the \(SVT_V\) in chimpanzees is similar to that in humans; the chimpanzee \(SVT_V\) grew consistently to 4.7 cm at five years of age. Thus, the different changes in the proportions of the SVT between chimpanzees and humans after late infancy do not involve the growth of the \(SVT_V\), and result from the greater growth of the chimpanzee \(SVT_H\).

As for the growth of the \(SVT_V\), the laryngeal skeleton of chimpanzees descends rapidly relative to the hyoid in early infancy, but the hyoid descends only slightly relative to the palate in that period (Nishimura et al., 2003). Thereafter, the laryngeal skeleton continues to descend gradually, but the hyoid descends rapidly. This finding is demonstrated in Figure 5 and Appendix 2. The distance from the hyoid to the vocal fold (\(HB-VF\)) increases rapidly to 1.1 cm by the first year of age, whereas the distance from the hyoid to the palate (\(HB-PP\)) shows only a negligible change in the first year of life (Nishimura et al., 2003). After late infancy, the \(HB-VF\) measure continues to increase to 2.2 cm at five years of age, but the \(HB-PP\) increases steeply from 2.9 cm to 5.4 cm at that age. A growth-related change in the descent of the laryngeal skeleton relative to the hyoid, similar to that of chimpanzees, is also observed in humans. Although the human hyoid descends relative to the palate even in early infancy, it continues to descend and this descent is more rapid than the descent of the laryngeal skeleton relative to the hyoid after late infancy (D.E. Lieberman et al., 2001). During that period, the epiglottis continues to descend gradually from the velum in chimpanzees, as in humans, to lengthen the oropharyngeal part rostral to the laryngopharyngeal parts (Fig. 5c, Appendix 2). Thus, after late infancy in chimpanzees, the major descent of the larynx is also accounted for by the descent of the hyoid, as in humans, and it causes elongation of an oropharyngeal part.

Discussion

This study provides no evidence for the idea that the rapid descent of the larynx or hyoid arose uniquely in the human lineage to allow the configuration of the equally long \(SVT_H\) and \(SVT_V\) to develop and thus permit the acquisition of speech in this lineage (Negus, 1949; P. Lieberman, 1984; Crelin, 1987; Houghton, 1993; Laitman and Reidenberg, 1993; Nishimura, 2003, 2005; Nishimura et al., 2003). Some gross anatomical studies using a cross-sectional ontogenetic series of embalmed chimpanzee specimens have suggested a developmental lowering in the position of the larynx or hyoid in chimpanzees (Kelemen, 1948; Avril, 1963; Jordan, 1971), but it was unclear whether this was analogous to the descent of the larynx in humans.
humans. Our previous study (Nishimura et al., 2003) confirmed that chimpanzees show a laryngeal descent with changes in the relative proportion of the SVTH to SVTV during early infancy, but this is not identical to the phenomenon in humans. Another MRI study using a cross-sectional ontogenetic series of embalmed chimpanzee specimens (Nishimura, 2005) showed relative changes in the SVTH and SVTV similar to those observed in the present study. However, the low resolution of the cross-sectional samples allowed the conclusion to be drawn that the adult configuration of the chimpanzee SVT results, at least partly, from a lesser descent of the larynx compared with that in humans, i.e., a lack of any major descent of the hyoid (Nishimura, 2005). Here, our longitudinal MRI study confirms that this descent of the hyoid is shared with chimpanzees, despite differences in the developmental trajectory of the SVT configuration between humans and chimpanzees.

However, there are slight differences in the timing and rate of the hyoid descent in humans and chimpanzees. Although this issue must be resolved in the future using a larger sample, it might be caused by the development of the laryngeal air sac in chimpanzees. Whereas humans have no sac, chimpanzees have a huge sac extending superiorly from the bilateral laryngeal ventricles, between the hyoid and thyroid cartilage bilaterally, and expanding anteriorly to the ventral region of the neck (Kelemen, 1948; Negus, 1949; Starck and Schneider, 1960; Avril, 1963; Hayama, 1970; Hewitt et al., 2002). This sac developed to reach the dorsal aspect of the body of the hyoid at four months of age for all three subjects in the present study (a small dark pouch on Fig. 1b; Nishimura et al., in press), and the associated anatomical changes might thereby contribute to elevate the hyoid in early infancy.

Regardless of these slight differences, the descent of the chimpanzee larynx and hyoid continues to lower the epiglottis from the velum and pulls the tongue down in the pharynx, elongating an oropharyngeal region facing the dorsal surface of the tongue. The chimpanzee epiglottis is believed to maintain contact with or to remain close to the velum even in adults, based on examinations using embalmed cadavers (Laitman and Heimbuch, 1982; P. Lieberman, 1984; Fitch, 2000a; Nishimura, 2005). However, MRI studies in living chimpanzees showed that the epiglottis loses contact with the velum in early infancy (Nishimura et al., 2003), as in humans (Sasaki et al., 1977), so that an oropharyngeal region is secured in the adult (Nishimura, 2006). Inevitably, the position of the epiglottis relative to the velum is affected by embalming artifacts (Nishimura, 2005) and varies actively in living animals (see Fig. 3; Fitch, 2000b). The latter suggests that the head posture at scanning (Nishimura et al., 2003; Nishimura, 2006; present study) might always lower or additionally elevate the epiglottis of the living subjects, although the present analysis carefully excluded any measurements in the nonresting subject (Ayumu at 52 months of age). However, in these studies, the neck of all the subjects was always bent slightly at the level below the vocal folds (Nishimura et al., 2003; Nishimura, 2006; present study). In addition, a varied head posture has little influence on the SVTV dimension, which reflects the position of the vocal folds relative to the palatal plane (Nishimura, 2005). This suggests that the position of the epiglottis, which is attached to the thyroid cartilage of the laryngeal skeleton, is also retained relative to the velum, regardless of the varied head posture. These facts suggest that varying head postures have limited influence on the over- or underestimation of oropharyngeal measures. This study showed the steady elongation of the oropharyngeal region, under the same conditions; unfortunately, there is no comparable information about the growth rates of the human oropharyngeal and laryngopharyngeal components. On the other hand, the human pharynx, defined as a linear dimension from VL to VF, increases from 2.63 cm at two to four years of age to 3.92 cm at nine to ten years of age, an increase of 49% (Fitch and Giedd, 1999). This increase is similar to the increases in sum of the oropharyngeal and laryngopharyngeal parts in the corresponding periods in chimpanzees: the sum of the two parts increases from 2.84 cm at two years of age to 4.30 cm at age five, an increase of 51% (Appendix 2). In humans, the pharyngeal dimension continues to increase to 5.64 cm in the adult (19 to 25 years of age), an increase of 114%. Despite little comparable information in living chimpanzees, an adult male subject at 23 years of age shows the sum of 5.58 cm, an increase of 96% (unpublished data, although the MR image is shown in Nishimura, 2006). Taking consideration of the secondary descent of the human larynx in the pubertal period (an additional elongation of the pharyngeal dimension; Fitch and Giedd, 1999), this indicates that the development of the pharyngeal dimension is similar in the two species. In addition, the timing in the loss of contact between the epiglottis and velum is similar approximately analogous in humans and chimpanzees: before 12 to 18 months in humans (Sasaki et al., 1977), and before 6 months in chimpanzees (Nishimura et al., 2003). This suggests that the development of the oropharyngeal dimension in chimpanzees is approximately analogous to that in humans. Although this issue must be resolved in the future using larger samples of humans and chimpanzees with varied head positions, it is probable that the major descent of the larynx, elongating an oropharyngeal region, was shared fully by the last common ancestor of chimpanzees and humans.

Uncertainty remains regarding the evolutionary path involving the major descent of the laryngeal skeleton and hyoid before the divergence of the human from the chimpanzee lineage. A study using computed tomography (CT) demonstrated a slight descent of the larynx and hyoid relative to the palate in macaques from birth to 24 years of age (Flügel and Rohen, 1991). The measurements reported by those authors cannot be directly compared with the studies of humans (D.E. Lieberman et al., 2001) and chimpanzees (Nishimura et al., 2003; Nishimura, 2005; present study) because of a discrepancy in the definitions. However, they suggest that an increase of the vertical dimension of the SVT in the first six months of macaques may be roughly similar to those in the corresponding periods of the other two species, although the measurements after six months of macaques must be excluded because of their
scarcity and irregularity. Based on the studies of dental eruption ages in macaques (Schultz, 1933; Iwamoto et al., 1984), the period from birth to six months of age corresponds to “early infancy” defined here for humans and chimpanzees. In macaques, the distance from the palate to the glottis increases from 1.2 cm on day one to 1.9 cm at six months of age, an increase of 58% (Flügel and Rohen, 1991). The SVTV in humans increases from 3.40 cm at one month to 4.92 cm at two years and nine months of age, an increase of 44% (D.E. Lieberman et al., 2001). In chimpanzees, it grows from 2.47 cm at four months to 3.26 cm at one year of age, an increase of 32% (Nishimura et al., 2003; see Appendix 2). Taking consideration of a lack of measurements in the periods earlier than one month for humans and four months for chimpanzees, these increases in the vertical dimension of the SVT in early infancy are similar in all three species. However, it must be noted that macaques have a laryngeal air sac opening above the anterior commissure of the glottis, extending superiorly to reach the dorsal aspect of the hyoid body and expanding inferiorly to the ventral region of the neck (Negus, 1949; Hayama, 1970; Hewitt et al., 2002). Although the development of this macaque air sac is still unclear, the developing sac might lower the laryngeal skeleton additionally relative to the hyoid in early infancy. In fact, although the descent of the larynx also continues after late infancy in macaques, CT scans show that the descent in macaques does not separate the epiglottis from the velum, producing no or little oropharyngeal space even in adults (Flügel and Rohen, 1991). Moreover, the laryngeal skeleton is separated from the hyoid and assured of mobility independent of the hyoid in hominoids, including humans, but this is in contrast to other anthropoids whose laryngeal skeletons are locked into the hyoid body and are tied tightly with or directly articulated with the greater horn of the hyoid, according to the comparative anatomy of adult cadaver specimens (Nishimura, 2003). Although this issue must be resolved in the future using larger samples and similar methods, these facts suggest that the descent of the larynx in macaques is not analogous to that seen in humans and chimpanzees. The major descent of the larynx may thus have evolved in two steps: the descent of the laryngeal skeleton relative to the hyoid in a common ancestor of all extant hominoids (Nishimura, 2003; Nishimura et al., 2003), and descent of the hyoid relative to the palate in a common ancestor of humans and chimpanzees. On the other hand, the descent of the laryngeal skeleton may have accompanied the descent of the hyoid. The hyoid provides the base of the tongue muscles and it is physically linked to the laryngeal skeleton, mandible, and cranial base through muscles, ligaments, and other soft tissues (Zemlin, 1988; Titze, 1994). The hyolaryngeal complex plays an important role in integrated functions of the pharynx and larynx, such as deglution, breathing, and vocalization (Negus, 1949; P. Lieberman, 1984; Crelin, 1987; Zemlin, 1988; Hiiemae and Palmer, 1999). The complexity of the anatomy and the integrated nature of the vital functions of the hyolaryngeal complex raise the possibility that the major descent of the laryngeal skeleton and hyoid may have arisen in a single shift to facilitate these integrated functions in a common ancestor of extant hominoids.

It is likely that the major descent of the larynx originally conferred advantages for functions unrelated to the sophisticated articulation of speech, because the faculty arose in the human lineage after its divergence from the chimpanzee lineage. This argument is also supported by the situation of a descended larynx in nonprimate mammals (Fitch, 2000b; Fitch and Reby, 2001; Weissengruber et al., 2002; Reby et al., 2005). Fitch and Reby (2001) showed that the position of the larynx in the males of some species of deer is similar to that in adult humans and that they lower the larynx to produce resonance at lower frequencies during the roaring of the rut season. This modification in vocalization possibly contributes to exaggeration of their perceived body size to increase the repulsive effects against other males (Reby et al., 2005). The ability to lower the hyolaryngeal complex is also found in the vocalizing of dogs, which have a resting anatomy with the epiglottis maintaining contact with the velum (Fitch, 2000b). That such dynamic lowering of the larynx exists in independent lineages clearly indicates the selective advantage or advantages of vocalization, independent of speech and language, and can account for the static low position of the larynx that was achieved with the developmental descent of the larynx in humans and chimpanzees. Such nonlinguistic advantage might be achieved by the secondary descent of the human larynx, which effects formant frequencies and probably underlies the vocal remodeling process during the puberty of human males (Fitch and Giedd, 1999). However, the integrated nature of the function of the hyolaryngeal complex makes it unlikely that any single selective advantage involving vocalization can account for the major descent and resulting low position of the larynx. The integration of deglutition and breathing suggests that an advantage in such physiology unrelated to vocalization may partly account for it (Nishimura, 2003). Such dynamic capabilities of the vocal tract and hyolaryngeal complex strongly indicate that critical evidence for this issue will be provided by evaluations of dynamic activities to perform a physiological function, not just by evaluations of the static anatomy of the SVT. Chimpanzees also show some dynamic activities, for example contraction of the tongue to pull up the hyoid (see Fig. 3) and retraction of the tongue to lower the hyolaryngeal complex (see Fig. 3 in Nishimura, 2005); unfortunately, there is no information about the relationships between such activities and physiological performances. Thus, such information in nonhuman primates and mammals will facilitate our understanding of the evolutionary pathway involving the major descent and static low position of the larynx against both phylogenetic and functional backgrounds.

Although this study suggested that the descent of the larynx and hyoid in chimpanzees is analogous to that in humans, but not to that in macaques, the adult configuration of the SVT in chimpanzees resembles that of macaques.
rather more than humans (Nishimura, 2005, 2006). Here, the ratio of SVTH and SVTV lengths in chimpanzees is higher than that in humans already in infancy, and it increases after late infancy—in contrast to humans—to result in a distinct final SVT configuration. Such differences between humans and chimpanzees were probably established by the evolution of facial flattening—the reductions of facial prognathism and projection—in the human lineage, but not by that of laryngeal descent itself. Similar developments in the pharyngeal configuration strongly suggest that the SVTV dimension in chimpanzees probably corresponds approximately to that in humans even at birth and later, despite smaller absolute values of the SVTV length in chimpanzees than in humans. This is supported in particular by the similar timing of the loss of contact between the epiglottis and velum. In contrast, the MR images show that the face is already projected and prognathic in chimpanzee infants (Fig. 1b). These characteristic differences in facial shape between the two species partly arise in prenatal development (Enlow, 1990; Ackermann and Krovitz, 2002; Penin et al., 2002; Mitteroecker et al., 2004). These facts suggest that the SVTH dimension in chimpanzees does not correspond to that in humans already at birth, although the absolute values of the SVTH length per se are similar. If so, this difference is principally involved in causing the higher ratio of the SVTH and SVTV lengths during infancy in chimpanzees compared with humans. In addition, after late infancy, the disparity of the SVTH growth trajectory contributes to differentiate further the SVT configuration in chimpanzees from that in humans. This fact supports the idea that the evolutionary attainment of a 1:1 proportion of the SVTH and SVTV required facial flattening rather than the additional descent of the larynx in the human lineage. Such reorganization of the facial architecture probably involved modifications to mandibular form. In anatomical terms, modification of the mandibular height affects the position of the hyoid (D.E. Lieberman et al., 2001), and therefore the facial flattening might have been accompanied by an additional descent of the hyoid. However, this does not mean that such facial flattening squeezed the tongue into the pharynx, which makes the larynx descend along the neck (Negus, 1949; Flügel and Rohen, 1991; Aiello, 1996, 1998; Mithen, 2005). The present study provides critical evidence for the idea that facial flattening merely produced a short, horizontal dimension for the tongue, fitting the SVTH configuration in the human lineage, independently of the descent of the hyoid, which had already increased the vertical dimension. These reductions in facial prognathism and projection were long-term trends in primate and human evolution (Klein, 1989; Wood, 1992; Aiello and Dean, 1998; Fleagle, 1999; Wood and Collard, 1999; Trinkaus, 2003), and therefore it is unlikely that the selective advantage afforded by speech originally accounted for this rearrangement. Thus, facial flattening has probably evolved only secondarily, with anatomical consequences that facilitated the evolution of sophisticated spoken language in the human lineage.

Conclusions

The major descent of the larynx and the reduced growth of the face essentially constitute the “ontogeny” of the morphological foundations of speech in humans (Negus, 1949; P. Lieberman, 1984; Crelin, 1987; Zemlin, 1988; Flügel and Rohen, 1991; Laitman and Reidenberg, 1993; Fitch and Giedd, 1999; D.E. Lieberman et al., 2001). However, this study does not support the “evolutionary” hypothesis that this descent occurred in the human lineage, and thereby produced the SVT configuration with an equally long SVTH and SVTV, leading to the origins of speech. The major descent of the larynx arose before the diversification of human and chimpanzee lineages, but it has not led to such a configuration of the SVT. Briefly, the term “descent” of the larynx has no implication for the proportional changes toward such an SVT configuration. The attainment of the SVT configuration unique to humans required facial flattening—modifications in the facial development to reduce the oral horizontal dimension—rather than additional laryngeal descent in the human lineage. These two separate developmental mechanisms probably arose independently during the evolution of nonhuman primates and humans, which were secondarily advantageous for speech in the latter. Therefore, it is unlikely that the faculty for speech evolved in humans through a single coordinated adaptive shift in all the relevant biological foundations. Instead, we propose the concept of a multiple-step evolution of the disparate elements in response to different selection pressures during the long period of primate evolution. This is also possibly the case for the evolution of the cognitive foundations of language (e.g., Matsuzawa, 2001). Rather than using one unique feature as evidence for the evolution of speech and language, a multidisciplinary approach should contribute greatly to our understanding of the evolution of the biological foundations of language.

Acknowledgments

We are grateful to A. Kato, K. Kumazaki, N. Maeda, S. Goto, K. Matsubayashi, M. Tomonaga, M. Tanaka, Y. Hamada, and the staff of PRI for daily veterinary care for the three chimpanzees and/or support for the MRI examinations. We thank D.E. Lieberman and his colleagues for kindly permitting us to use the numerical data from D.E. Lieberman et al. (2001). Portions of this work were performed at PRI under 2004 and 2005 PRI Cooperative Research Programs (Grant 4-1 and 3-1 to T. N.). This work was supported by a Research Fellowship of the Japan Society for the Promotion of Science (JSPS) for Young Scientist (Grant 16000326 to T. N.), by Grants-in-Aid for the Biodiversity Research of the 21st Century Center of Excellence (Grant A14 to Kyoto University), and for Specially Promoted Research (Grants 12002009, 16002001 to T. M.) from the Ministry of Education, Culture, Sports, Sciences, and Technology of Japan, and by the JSPS Core-to-Core Program HOPE (Grant 15001 to T.M.)
Appendix 1. Subjects used, their ages at the times of scans, and major parameters of scanning

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Age (months)</th>
<th>Pixel sizes(^1) (mm/pixel)</th>
<th>FOV (mm)</th>
<th>Thickness (inter-slice gap)(^2) (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ayumu</td>
<td>4</td>
<td>0.74</td>
<td>190</td>
<td>2.7 (0.8)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.70</td>
<td>180</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.74</td>
<td>190</td>
<td>2.7 (0.8)</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.70</td>
<td>180</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>0.70</td>
<td>180</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>0.90</td>
<td>230</td>
<td>2.7 (0.8)</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.86</td>
<td>220</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.98</td>
<td>250</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>1.09</td>
<td>280</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>0.98</td>
<td>250</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>0.98</td>
<td>250</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>0.98</td>
<td>250</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>54</td>
<td>0.98</td>
<td>250</td>
<td>2.7 (0.8)</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.98</td>
<td>250</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td>Cleo</td>
<td>4</td>
<td>0.70</td>
<td>180</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.70</td>
<td>180</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.70</td>
<td>180</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.74</td>
<td>190</td>
<td>2.7 (0.8)</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>0.70</td>
<td>180</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.86</td>
<td>220</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>0.98</td>
<td>250</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>0.98</td>
<td>250</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>0.98</td>
<td>250</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>0.98</td>
<td>250</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>54</td>
<td>0.98</td>
<td>250</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.98</td>
<td>250</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td>Pal</td>
<td>4</td>
<td>0.70</td>
<td>180</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.74</td>
<td>190</td>
<td>2.7 (0.8)</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.70</td>
<td>180</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.70</td>
<td>180</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>0.70</td>
<td>180</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>0.86</td>
<td>220</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.98</td>
<td>250</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>0.98</td>
<td>250</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>0.98</td>
<td>250</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>0.98</td>
<td>250</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>54</td>
<td>0.98</td>
<td>250</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.98</td>
<td>250</td>
<td>3.0 (0.5)</td>
</tr>
</tbody>
</table>

\(^1\) The values varied irregularly during infancy, but had little influence on the accuracy in measurements (see Results).

\(^2\) This means that the slice interval of all scans was 3.5 mm, regardless of differences in slice thickness and inter-slice gap.

Appendix 2. Measurements of dimensions

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Age (months)</th>
<th>SVT(_H)</th>
<th>SVT(_V)</th>
<th>Oropharynx</th>
<th>Laryngopharynx</th>
<th>HB— (op)</th>
<th>HB— (lp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ayumu</td>
<td>9</td>
<td>5.93</td>
<td>2.75</td>
<td>0.29</td>
<td>1.60</td>
<td>3.00</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>6.33</td>
<td>3.41</td>
<td>0.49</td>
<td>1.78</td>
<td>2.98</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>6.59</td>
<td>3.53</td>
<td>0.67</td>
<td>2.07</td>
<td>3.39</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>6.78</td>
<td>3.84</td>
<td>0.89</td>
<td>1.97</td>
<td>3.71</td>
<td>1.28</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>7.17</td>
<td>3.87</td>
<td>1.12</td>
<td>2.05</td>
<td>4.30</td>
<td>1.41</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>8.03</td>
<td>4.25</td>
<td>1.23</td>
<td>2.08</td>
<td>4.46</td>
<td>1.42</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>8.80</td>
<td>4.47</td>
<td>1.35</td>
<td>2.38</td>
<td>4.68</td>
<td>1.58</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>9.04</td>
<td>4.65</td>
<td>1.45</td>
<td>2.54</td>
<td>4.74</td>
<td>1.42</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>9.25</td>
<td>4.67</td>
<td>1.48</td>
<td>2.61</td>
<td>5.06</td>
<td>1.38</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>9.06</td>
<td>4.63</td>
<td>1.11</td>
<td>2.85</td>
<td>4.17</td>
<td>1.61</td>
</tr>
<tr>
<td></td>
<td>54</td>
<td>9.63</td>
<td>4.90</td>
<td>1.48</td>
<td>2.86</td>
<td>5.50</td>
<td>1.95</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>10.16</td>
<td>5.08</td>
<td>1.66</td>
<td>3.03</td>
<td>5.57</td>
<td>2.19</td>
</tr>
</tbody>
</table>

References


