2. Physiological Processes of Speech Production

K. Honda

Speech sound is a wave of air that originates from complex actions of the human body, supported by three functional units: generation of air pressure, regulation of vibration, and control of resonators. The lung air pressure for speech results from functions of the respiratory system during a prolonged phase of expiration after a short inhalation. Vibrations of air for voiced sounds are introduced by the vocal folds in the larynx; they are controlled by a set of laryngeal muscles and airflow from the lungs. The oscillation of the vocal folds converts the expiratory air into intermittent airflow pulses that result in a buzzing sound. The narrow constrictions of the airway along the tract above the larynx also generate transient source sounds; their pressure gives rise to an airstream with turbulence or burst sounds. The resonators are formed in the upper respiratory tract by the pharyngeal, oral, and nasal cavities. These cavities act as resonance chambers to transform the laryngeal buzz or turbulence sounds into the sounds with special linguistic functions. The main articulators are the tongue, lower jaw, lips, and velum. They generate patterned movements to alter the resonance characteristics of the supra-laryngeal airway. In this chapter, contemporary views on phonatory and articulatory mechanisms are summarized to illustrate the physiological processes of speech production, with brief notes on their observation techniques.

2.1 Overview of Speech Apparatus

The speech production apparatus is a part of the motor system for respiration and alimentation. The form of the system can be characterized, when compared with those of other primates, by several unique features, such as small red lips, flat face, compact teeth, short oral cavity with a round tongue, and long pharynx with a low larynx position. The functions of the system are also uniquely advanced by the developed brain with the language areas, direct neural connections from the cortex to motor nuclei, and dense neural supply to each muscle. Independent control over phonation and articulation is a human-specific ability. These morphological and neural changes along human evolution reorganized the original functions of each component into an integrated motor system for speech communication.

The speech apparatus is divided into the organs of phonation (voice production) and articulation (settings of the speech organs). The phonatory organs (lungs and larynx) create voice source sounds by setting the driving air pressure in the lungs and parameters for vocal fold vibration at the larynx. The two organs together adjust the pitch, loudness, and quality of the voice, and further generate prosodic patterns of speech. The articulatory organs give resonances or modulations to the voice source and generate additional sounds for some consonants. They consist of the lower jaw, tongue,

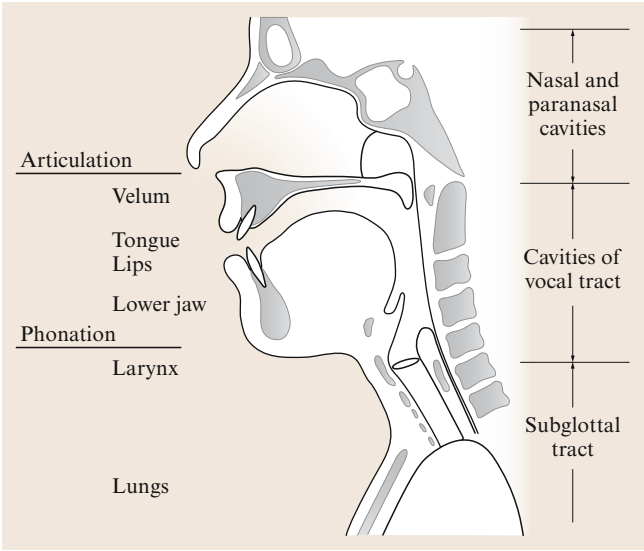

Fig. 2.1 Sketch of a speech production system. Physiological processes of speech production are realized by combined sequential actions of the speech organs for phonation and articulation. These activities result in sound propagation phenomena at the three levels: subglottal cavities, cavities of the vocal tract, and nasal and paranasal cavities

lips, and the velum. The larynx also takes a part in the articulation of voiced/voiceless distinctions. The tongue and lower lip attach to the lower jaw, while the velum is loosely combined with other articulators. The constrictor muscles of the pharynx and larynx also participate in articulation as well as in voice quality control. The phonatory and articulatory systems influence each other mutually, while changing the vocal tract shape for producing vowels and consonants. Figure 2.1 shows a schematic drawing of the speech production system.

2.2 Voice Production Mechanisms

Generation of voice source requires adequate configuration of the airflow from the lungs and vocal fold parameters for oscillation. The sources for voiced sounds are the airflow pulses generated at the larynx, while those for some consonants (i. e., stops and fricatives) are airflow noises made at a narrow constriction in the vocal tract. The expiratory and inspiratory muscles together regulate relatively constant pressure during speech. The laryngeal muscles adjust the onset/offset, amplitude, and frequency of vocal fold vibration.

2.2.1 Regulation of Respiration

The respiratory system is divided into two segments: the conduction airways for ventilation between the atmosphere and the lungs, and the respiratory tissue of the lungs for gas exchange. Ventilation (i. e., expiration and inhalation) is carried out by movements of the thorax, diaphragm, and abdomen. These movements involve actions of respiratory muscles and elastic recoil forces of the system. During quiet breathing, the lungs expand to inhale air by the actions of inspiratory muscles (diaphragm, external intercostal, etc.), and expel air by the elastic recoil force of the lung tissue, diaphragm, and cavities of the thorax and abdomen. In effort expiration, the expiratory muscles (internal intercostals, abdominal muscles, etc.) come into action. The inspiratory and expiratory muscles work alternately, making the thorax expand and contract during deep breathing.

During speech production, the respiratory pattern changes to a longer expiratory phase with a shorter inspiratory phase during quiet breathing. Figure 2.2 shows a conventional view of the respiratory pattern during

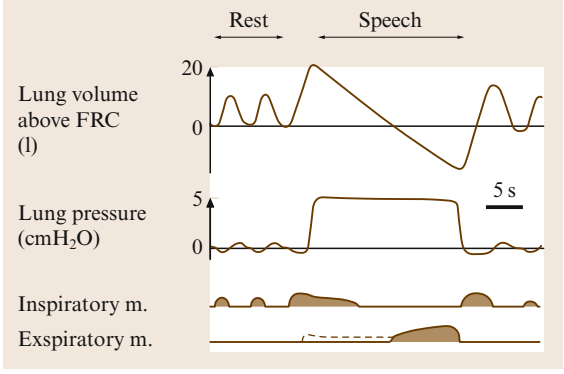

Fig. 2.2 Respiratory pattern during speech. Top two curves show the changes in the volume and pressure in the lungs. The bottom two curves show schematic activity patterns of the inspiratory and expiratory muscles (after [2.1]). The dashed line for the expiratory muscles indicates their predicted activity for expiration

speech [2.1]. The thorax is expanded by inspiration prior to initiation of speech, and then compressed by elastic recoil force by the tissues of the respiratory system to the level of the functional residual capacity (FRC). The lung pressure during speech is kept nearly constant except for the tendency of utterance initial rise and final lowering. In natural speech, stress and emphasis add local pressure increases. The constant lung pressure is due to the actions of the inspiratory muscles to prevent excessive airflow and maintain the long expiratory phase. As speech continues, the lung volume decreases gradually below the level of FRC, and the lung pressure is then maintained by the actions of the expiratory muscles that actively expel air from the lung. It has been argued whether the initiation of speech involves only the elastic recoil forces of the thorax to generate expiratory airflow. Indeed, a few studies have suggested that not only the thoracic system but also the abdominal system assists the regulation of expiration during speech [2.2, 3], as shown by the dashed line in Fig. 2.2. Thus, the contemporary view of speech respiration emphasizes that expiration of air during speech is not a passive process but a controlled one with co-activation of the inspiratory and expiratory muscles.

2.2.2 Structure of the Larynx

The larynx is a small cervical organ located at the top of the trachea making a junction to the pharyngeal cavity: it primarily functions to prevent foreign material from entering the lungs. The larynx contains several rigid structures such as the cricoid, thyroid, arytenoid, epiglottic, and other smaller cartilages. Figure 2.3a shows the arrangement of the major cartilages and the hyoid bone. The cricoid cartilage is ring-shaped and supports the lumen of the laryngeal cavity. It offers two bilateral articulations to the thyroid and arytenoid cartilages at the cricothyroid and cricoarytenoid joints, respectively. The thyroid cartilage is a shield-like structure that offers attachments to the vocal folds and the vestibular folds. The arytenoid cartilages are bilateral tetrahedral cartilages that change in location and orientation between phonation and respiration. The whole larynx is mechanically suspended from the hyoid bone by muscles and ligaments.

The gap between the free edges of the vocal folds is called the *glottis*. The space is divided into two portions by the vocal processes of the arytenoid cartilages: the membranous portion in front (essential for vibration) and cartilaginous portion in back (essential for respiration). The glottis changes its form in various ways during speech: it narrows by adduction and widens by abduction of the vocal folds. Figure 2.3b shows that this movement is carried out by the actions of the intrinsic laryngeal muscles that attach to the arytenoid cartilages. These muscles are functionally divided into the adductor and abductor muscles. The adductor muscles include the thyroarytenoid muscles, lateral cricoarytenoid, and arytenoid muscles, and the abductor muscle is the posterior cricoarytenoid muscle. The glottis also changes in length according to the length of the vocal folds, which takes place mainly at the membranous portion. The length of the glottis shows a large developmental sexual variation. The membranous length on average is 10 mm in adult females and 16 mm in adult males, while the cartilaginous length is about 3 mm for both [2.4].

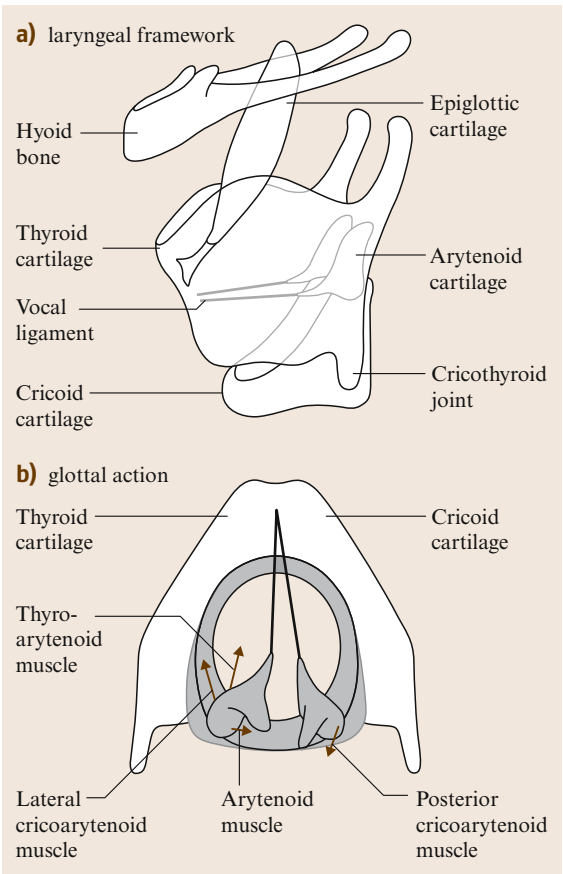

Fig. 2.3a,b Laryngeal framework and internal structures. **(a)** Oblique view of the laryngeal framework, which includes the hyoid bone and four major cartilages. **(b)** Adduction (*left*) and abduction (*right*) of the glottis and the effects of the intrinsic laryngeal muscles

2.2.3 Vocal Fold and its Oscillation

The larynx includes several structures such as the subglottic dome, vocal folds, ventricles, vestibular folds, epiglottis, and aryepiglottic folds, as shown in Fig. 2.4a. The vocal folds run anteroposteriorly from the vocal processes of the arytenoid cartilages to the internal surface of the thyroid cartilage. The vocal fold tissue consists of the thyroarytenoid muscle, vocal ligament, lamina propria, and mucous membrane. They form a special layer structure that yields to aerodynamic forces to oscillate, which is often described as the *body-cover* structure [2.5].

During voiced speech sounds, the vocal folds are set into vibration by pressurized air passing through the membranous portion of the narrowed glottis. The glottal airflow thus generated induces wave-like motion of the vocal fold membrane, which appears to propagate from the bottom to the top of the vocal fold edges. When this oscillatory motion builds up, the vocal fold membranes on either side come into contact with each other, resulting in repetitive closing and opening of the glottis. Figure 2.4b shows that vocal fold vibration repeats four phases within a cycle: the closed phase, opening phase, open phase, and closing phase. The conditions that determine vocal fold vibration are the stiffness and mass of the vocal folds, the width of the glottis, and the pressure difference across the glottis.

The aerodynamic parameters that regulate vocal fold vibration are the transglottal pressure difference and glottal airflow. The former coincides with the measure of subglottal pressure during mid and low vowels, which is about 5–10 cm H_2O in comfortable loudness and pitch (1 cm H_2O = 0.98 hPa). The latter also coincides with the average measure of oral

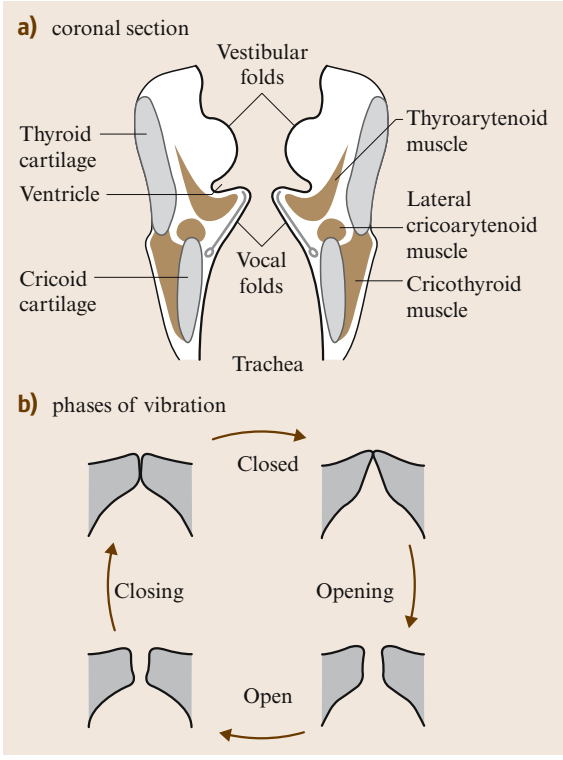

Fig. 2.4a,b Vocal folds and their vibration pattern. **(a)** Coronal section of the larynx, showing the tissues of the vocal and vestibular (false) folds. The cavity of the larynx includes supraglottic and subglottic regions. **(b)** Vocal-fold vibration pattern and glottal shapes in open phases. As the vocal-fold edge deforms in a glottal cycle, the glottis follows four phases: closed, opening, open and closing

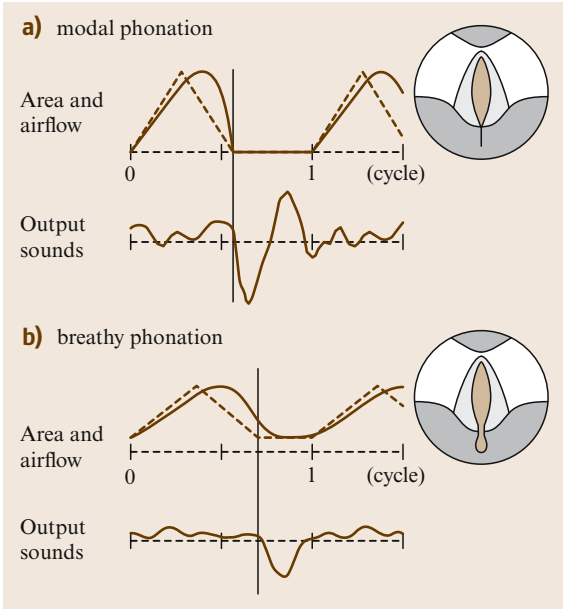

Fig. 2.5a,b Changes in glottal area and airflow in relation to output sounds during 1.5 glottal cycles from glottal opening, with glottal shapes at peak opening (in the circles). **(a)** In modal phonation with complete glottal closure in the closed phase, glottal closure causes abrupt shut-off of glottal airflow and strong excitation of the air in the vocal tract during the closed phase. **(b)** In breathy phonation, the glottal closure is incomplete, and the airflow wave includes a DC component, which results in weak excitation of the tract

airflow during vowel production, which is roughly 0.1–0.2 l/s. These values show a large individual variation: the pressure range is 4.2–9.6 cm H_2O in males and 4.4–7.6 cm H_2O in females, while the airflow rate ranges between 0.1–0.3 l/s in males and 0.09–0.21 l/s in females [2.6].

Figure 2.5 shows schematically the relationship between the glottal cycle and volumic airflow change in normal and breathy phonation. The airflow varies within each glottal cycle, reflecting the cyclic variation of the glottal area and subglottal pressure. The glottal area curve roughly shows a triangular pattern, while the airflow curve shows a skew of the peak to the right due to the inertia of the air mass within the glottis [2.7]. The closure of the glottis causes a discontinuous decrease of the glottal airflow to zero, which contributes the main source of vocal tract excitation, as shown in Fig. 2.5a. When the glottal closure is more abrupt, the output sounds are more intense with richer harmonic components [2.8]. When the glottal closure is incomplete in soft and breathy voices or the cartilaginous portion of the glottis is open to show the *glottal chink*, the airflow includes a direct-current (DC) component and exhibits a gradual decrease of airflow, which results in a more sinusoidal waveform and a lower intensity of the output sounds, as shown in Fig. 2.5b.

Laryngeal control of the oscillatory patterns of the vocal folds is one of the major factors in voice quality control. In sharp voice, the open phase of the glottal cycle becomes shorter, while in soft voice, the open phase becomes longer. The ratio of the open phase within a glottal cycle is called the *open quotient* (OQ), and the ratio of the closing slope to the opening slope in the glottal cycle is called the *speed quotient* (SQ). These two parameters determine the slope of the spectral envelope. When the open phase is longer (high OQ) with a longer closing phase (low SQ), the glottal airflow becomes more sinusoidal, with weak harmonic components. Contrarily, when the open phase is shorter (low OQ), glottal airflow builds up to pulsating waves with rich harmonics. In modal voice, all the vocal fold layers are involved in vibration, and the membranous glottis is completely closed during the closed phase of each cycle. In falsetto, only the edges of the vocal folds vibrate, glottal closure becomes incomplete, and harmonic components reduce remarkably.

The oscillation of the vocal folds during natural speech is quasiperiodic, and cycle-to-cycle variation are observed in speech waveforms as two types of measures: *jitter* (frequency perturbation) and *shimmer* (amplitude perturbation). These irregularities appear to arise from combinations of biomechanical (vocal fold asymmetry), neurogenic (involuntary activities of laryngeal muscles), and aerodynamic (fluctuations of airflow and subglottal pressure) factors. In sustained phonation of normal voice, the jitter is about 1% in frequency, and the shimmer is about 6% in amplitude.

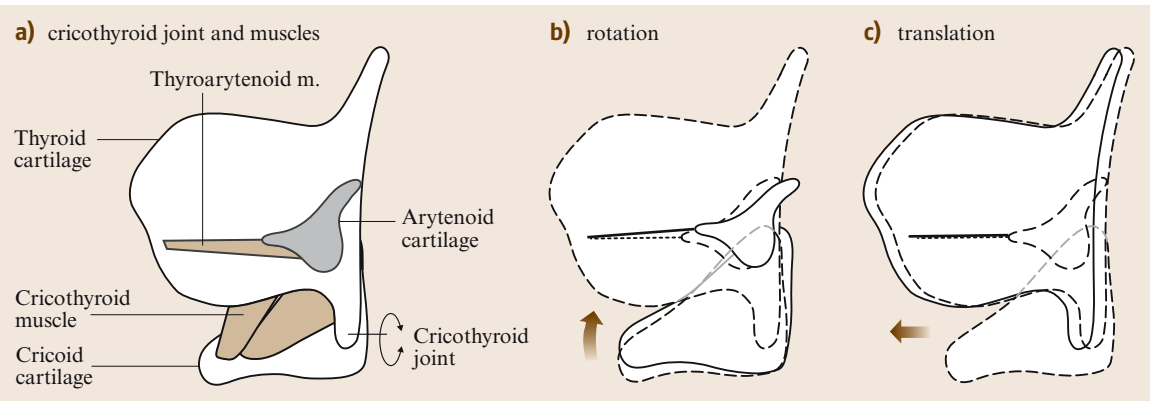

Fig. 2.6a–c Cricothyroid joint and F_0 regulation mechanism. **(a)** The cricothyroid joint is locally controlled by the thyroarytenoid and two parts of the cricothyroid muscles: Pars recta (anterior) and pars obliqua (posterior). As F_0 rises, the thyroid cartilage advances and cricoid cartilage rotates to the direction to stretch the vocal folds, which leads to the increases in the stiffness of vocal fold tissue and in the natural resonance frequency of the vocal folds. **(b)** Rotation of the cricothyroid joint is caused mainly by the action of the pars recta to raise the cricoid arch. **(c)** Translation of the joint is produced mainly by the pars obliqua

2.2.4 Regulation of Fundamental Frequency (F_0)

The fundamental frequency (F_0) of voice is the lowest harmonic component in voiced sounds, which conforms to the natural frequency of vocal fold vibration. F_0 changes depending on two factors: regulation of the length of the vocal folds and adjustment of aerodynamic factors that satisfy the conditions necessary for vocal fold vibration. In high F_0, the vocal folds become thinner and longer; while in low F_0, the vocal folds become shorter and thicker. As the vocal folds are stretched by separating their two attachments (the anterior commissure and vocal processes), the mass per unit length of the vocal fold tissue is reduced while the stiffness of the tissue layer involved in vibration increases. Thus, the mass is smaller and the stiffness is greater for higher F_0 than lower F_0, and it follows that the characteristic frequency of vibrating tissue increases for higher F_0. The length of the vocal folds is adjusted by relative movement of the cricoid and thyroid cartilages. Its natural length is a determinant factor of individual difference in F_0. The possible range of F_0 in adult speakers is about 80–400 Hz in males, and about 120–800 Hz in females.

The thyroid and cricoid cartilages are articulated at the cricothyroid joint. Any external forces applied to this joint cause rotation and translation (sliding) of the joint, which alters the length of the vocal folds. It is well known that the two joint actions are brought about by the contraction of the cricothyroid muscle to approximate the two cartilages at their front edges. Figure 2.6 shows two possible actions of the cricothyroid muscle on the joint: rotation by the pars recta and translation of the pars obliqua [2.9]. Questions still remain as to whether each part of the cricothyroid conducts pure actions of rotation or translation, and as to which part is more responsible for determining F_0.

The extrinsic laryngeal muscles can also apply external forces to this joint as a supplementary mechanism for regulating F_0 [2.10]. The most well known among the activities of the extrinsic muscles in this regulation is the transient action of the sternohyoid muscle observed as F_0 falls. Since this muscle pulls down the hyoid bone to lower the entire larynx, larynx lowering has long been thought to play a certain role in F_0 lowering. Figure 2.7 shows a possible mechanism of F_0 lowering by vertical larynx movement revealed by magnetic resonance imaging (MRI). As the cricoid cartilage descends along the anterior surface of the cervical spine, the cartilage rotates in a direction that shortens the vocal folds because the cervical spine shows anterior convexity at the level of the cricoid cartilage [2.11].

Aerodynamic conditions are an additional factor that alters F_0, as seen in the local rises of the subglottal pressure during speech at stress or emphasis. The increase of the subglottal air pressure results in a larger airflow rate and a wider opening of the glottis, which causes greater deformation of the vocal folds with larger average tissue stiffness. The rate of F_0 increase due to the subglottal pressure is reported to be about 2–5 Hz/cmH_2O when the chest cavity is compressed externally, and is observed to be 5–15 Hz/cmH_2O, when

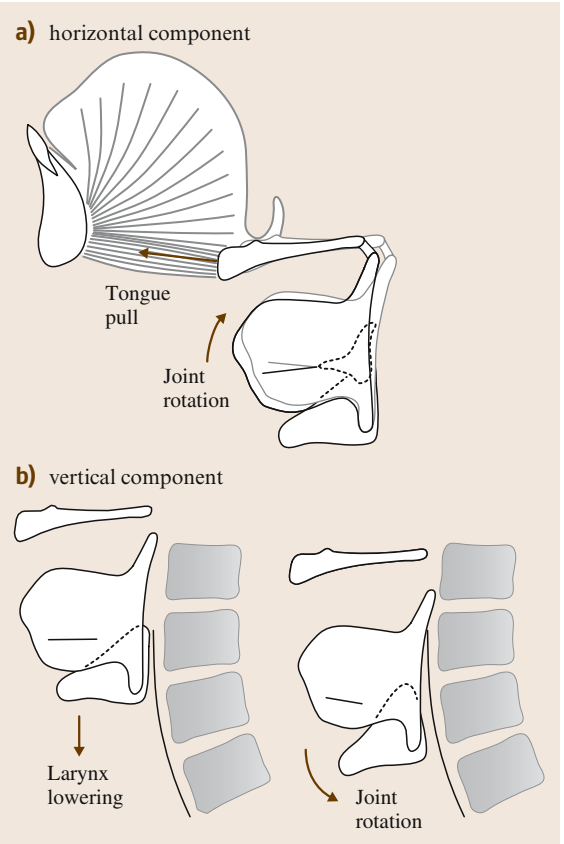

Fig. 2.7a,b Extrinsic control of F_0. Actions of the cricothyroid joint are determined not only by the cricothyroid muscle but also by other laryngeal muscles. Any external forces applied to the joint can activate the actions of the joint. **(a)** In F_0 raising, advancement of the hyoid bone possibly apply a force to rotate the thyroid cartilage. **(b)** In F_0 lowering, the cricoid cartilage rotates as its posterior plate descends along the anterior convexity of the cervical spine

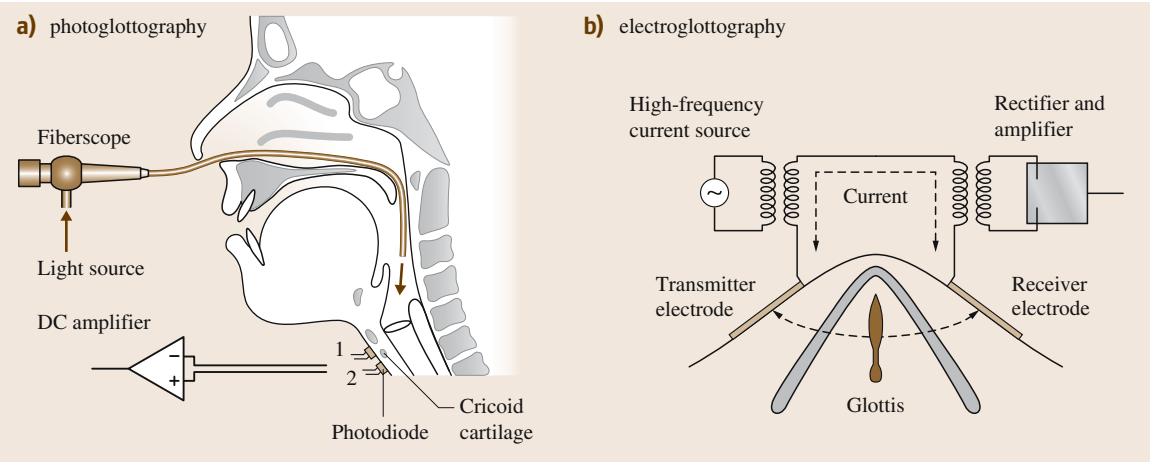

Fig. 2.8a,b Glottographic methods. **(a)** PGG with fiberscopy uses a photodetector attached near the cricothyroid cartilage in two locations: one attachment for measuring vibrations, and two attachment for glottal gestures. **(b)** EGG uses a pair of electrodes on the skin above the thyroid lamina to form a induction circuit to record electrical currents passed through the vocal-fold edges

it is measured between the beginning and end of speech utterances.

2.2.5 Methods for Measuring Voice Production

Speech production mechanisms arise from the functions of the internal organs of the human body that are mostly invisible. Therefore, better understanding of speech production processes relies on the development of observation techniques. The lung functions in speech can be assessed by the tools for aerodynamic measurements, while examination of the larynx functions during speech requires special techniques for imaging and signal recording.

Monitoring Respiratory Functions

Respiratory functions during speech are examined by recording aerodynamic measurements of lung volume, airflow, and pressure. Changes in lung volume are monitored with several types of plethysmography (e.g., whole-body, induction, and magnetic). The airflow from the mouth is measured with pneumotachography using a mask with pressure probes (differential-pressure anemometry) or thermal probes (hot-wire anemometry). Measurements of the subglottal pressure require a tracheal puncture of a needle with a pressure sensor or a thin catheter-type pressure transducer inserted from the nostril to the trachea via the cartilaginous part of the glottis.

Laryngeal Endoscopy

Imaging of the vocal folds during speech has been conducted with a combination of an endoscope and video camera. A solid-type endoscope is capable of observing vocal fold vibration with stroboscopic or real-time digital imaging techniques during sustained phonation. The flexible endoscope is beneficial for video recording of glottal movements during speech with a fiber optic bundle inserted into the pharynx through the nostril via the velopharyngeal port. Recently, an electronic type of flexible endoscope with a built-in image sensor has become available.

Glottography

Glottography is a technique to monitor vocal fold vibration as a waveform. Figure 2.8 shows two types of glottographic techniques. Photoglottography (PGG) detects light intensity modulated by the glottis using an optical sensor. The sensor is placed on the neck and a flexible endoscope is used as a light source. The signal from the sensor corresponds to the glottal aperture size, reflecting vocal fold vibration and glottal adduction–abduction movement. Electroglottography (EGG) records the contact of the left and right vocal fold edges during vibration. High-frequency current is applied to a pair of surface electrodes placed on the skin above the thyroid lamina, which detect a varying induction current that corresponds to the change in vocal fold contact area.

2.3 Articulatory Mechanisms

Speech articulation is the most complex motor activity in humans, producing concatenations of phonemes into syllables and syllables into words using movements of the speech organs. These articulatory processes are conducted within a phrase of a single expiratory phase with continuous changes of vocal fold vibration, which is one of the human-specific characteristics of sound production mechanisms.

2.3.1 Articulatory Organs

Articulatory organs are composed of the rigid organ of the lower jaw and soft-tissue organs of the tongue, lips, and velum, as illustrated in Fig. 2.9. These organs together alter the resonance of the vocal tract in various ways and generate sound sources for consonants in the vocal tract. The tongue is the most important articulatory organ, and changes the gross configuration of the vocal tract. Deformation of the whole tongue determines vowel quality and produces palatal, velar, and pharyngeal consonants. Movements of the tongue apex and blade contribute to the differentiation of dental and alveolar consonants and the realization of retroflex consonants. The lips deform the open end of the vocal tract by various types of gestures, assisting the production of vowels and labial consonants. Actions of these soft-tissue organs are essentially based on contractions of the muscles within these organs, and their mechanism is often compared with the *muscular hydrostat*. Since the tongue and lips have attachments to the lower jaw, they are interlocked with the jaw to open the mouth. The velum controls opening and closing of the velopharyngeal port, and allows distinction between nasal and oral sounds. Additionally, the constrictor muscles of the pharynx adjust the lateral width of the pharyngeal cavity, and their actions also assist articulation for vowels and back consonants.

Upper Jaw

The upper jaw, or the maxilla with the upper teeth, is the structure fixed to the skull, forming the palatal dome on the arch of the alveolar process with the teeth. It forms a fixed wall of the vocal tract and does not belong to the articulatory organs: yet it is a critical structure for speech articulation because it provides the frame of reference for many articulatory gestures. The structures of the upper jaw offer the location for contact or approximation by many parts of the tongue such as the apex, blade, and dorsum. The phonetics literature describes the place of articulation as classified according to the locations of lingual approximation along the upper jaw for dental, alveolar, and palatal consonants. The hard palate is covered by the thick mucoperiosteum, which has several transverse lines of mucosal folds called the *palatine rugae*.

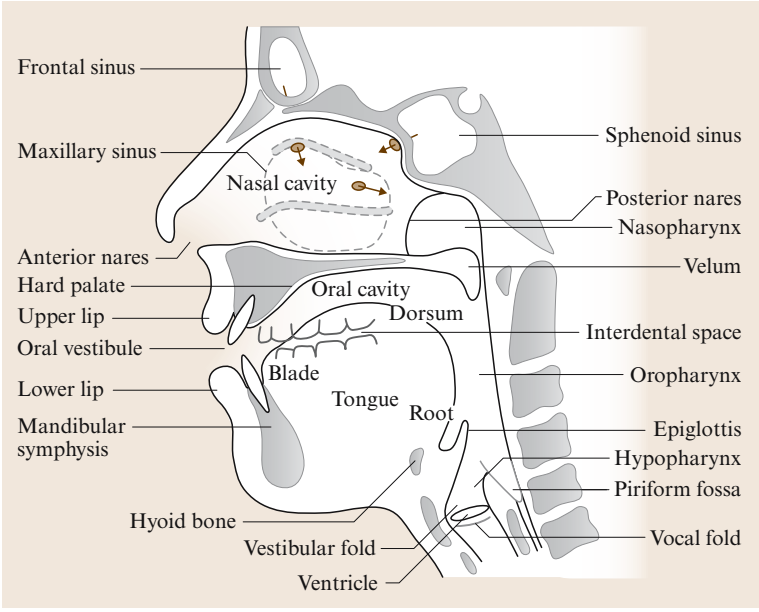

Fig. 2.9 Illustration of the articulatory system with names of articulators and cavities

Lower Jaw

The lower jaw, or the mandible with the lower teeth, is the largest rigid motor organ among the speech production apparatus. Its volume is about 100 cm^3. As well as playing the major role in opening and closing the mouth, it provides attachments for many speech muscles and supports the tongue, lips, and hyoid bone.

Figure 2.10 shows the action of the jaw and the muscles used in speech articulation. The mandible articulates with the temporal bone at the temporomandibular joint (TMJ) and brings about jaw opening–closing actions by rotation and translation. The muscles that control jaw movements are generally called the masticatory muscles. The jaw opening muscles are the digastric and lateral pterygoid muscles. The strap muscles, such as the geniohyoid and sternohyoid, also assist jaw opening. The jaw closing muscles include the masseter, temporalis, and medial pterygoid muscles. While the larger muscles play major roles in biting and chewing, comparatively small muscles are used for speech articulation. The medial pterygoid is mainly used for jaw closing in articulation, and the elastic recoil force of the connective tissues surrounding the mandible is another factor for closing the jaw from its open position.

Tongue

The tongue is an organ of complex musculature [2.12]. It consists of a round body occupying its main mass and a short blade with an apex. Its volume is approximately 100 cm^3, including the muscles in the tongue floor. The tongue body moves in the oral cavity by variously deforming its voluminous mass, while the tongue blade alters its shape and changes the angle of the tongue apex. Deformation of the tongue tissue is caused by contractions of the extrinsic and intrinsic tongue muscles, which are illustrated schematically in Fig. 2.11.

The extrinsic tongue muscles are those that arise outside of the tongue and end within the tongue tissue. This group includes four muscles, the genioglossus, hyoglossus, styloglossus, and palatoglossus muscles, although the former three muscles are thought to be involved in the articulation of the tongue. The palatoglossus muscle participates in the lowering of the velum as discussed later.

The genioglossus is the largest and strongest muscle in the tongue. It begins from the posterior aspect of the mandibular symphysis and runs along the midline of the tongue. Morphologically, it belongs to the triangular muscle, and its contraction effects differ across portions of the muscle. Therefore, the genioglossus is divided functionally into the anterior, middle, and posterior bundles. The anterior and middle bundles run midsagittally, and their contraction makes the midline groove of the tongue for the production of front vowels. The anterior bundle often makes a shallow notch on the tongue surface called the *lingual fossa* and assists elevation of the tongue apex. The middle bundle runs obliquely, and advances the tongue body for front vowels. The posterior bundle of the genioglossus runs midsagittally and

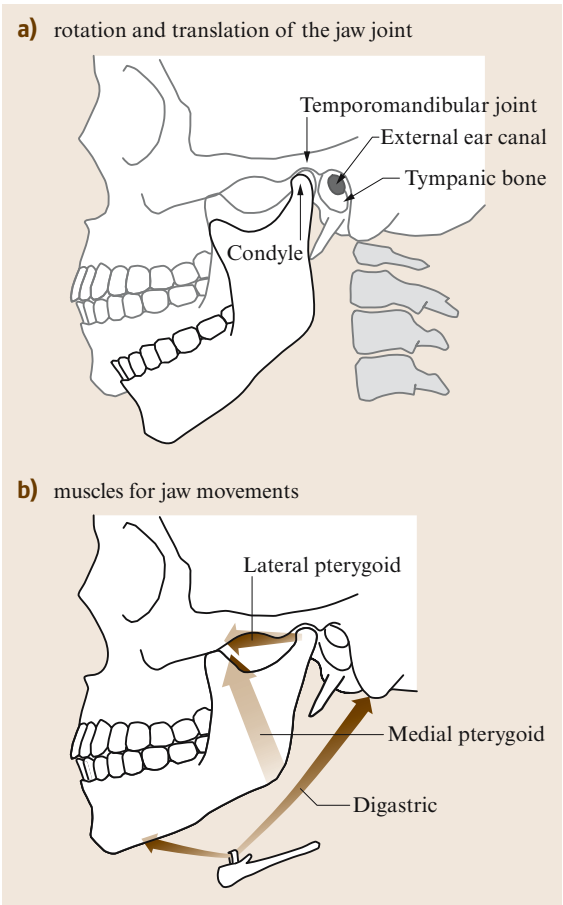

Fig. 2.10a,b Actions of the temporomandibular joint and muscles for jaw opening and closing. **(a)** The lower jaw opens by rotation and translation of the mandible at the temporomandibular joint. Jaw translation is needed for wide opening of the jaw because jaw rotation is limited by the narrow space between the condyle and tympanic bone. **(b)** Jaw opening in speech depends on the actions of the digastric and medial pterygoid muscles with support of the strap muscles. Jaw closing is carried out by the contraction of the lateral pterygoid muscle and elastic recoil forces of the tissues surrounding the jaw

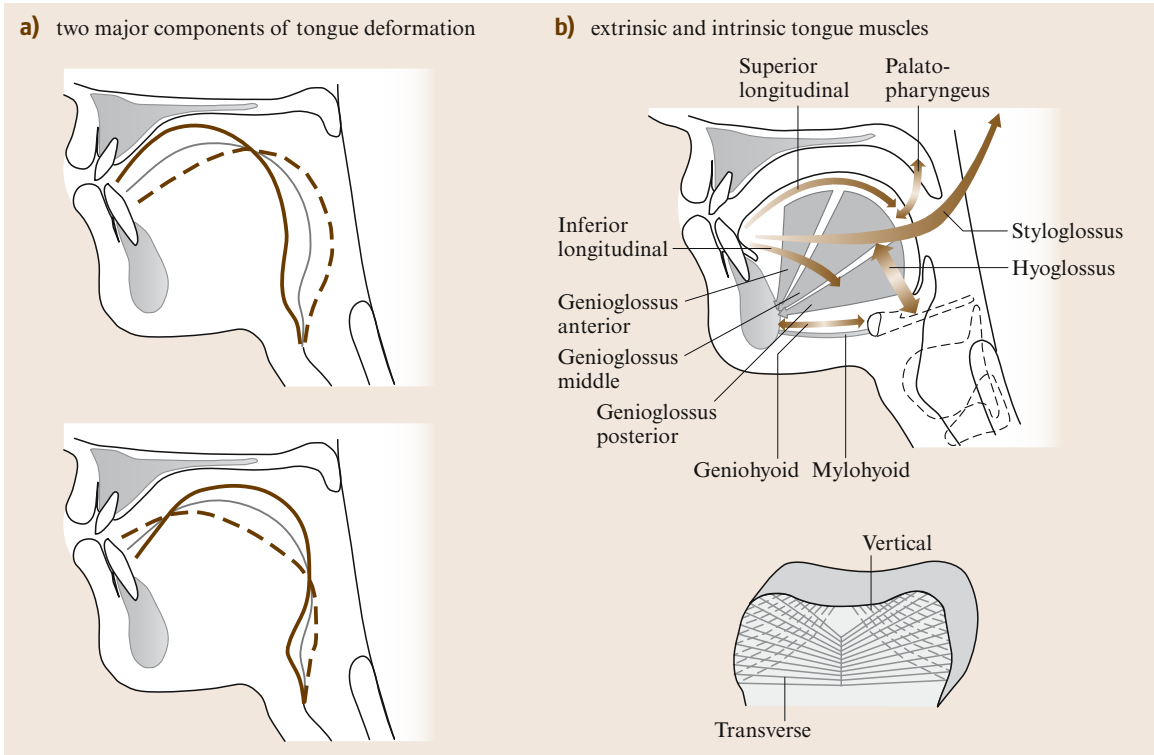

Fig. 2.11a,b Actions of the tongue and its musculature. **(a)** Major components of tongue deformation are high-front vs.low-back (*top*) and high back versus low front (*bottom*) motions, (after [2.14]). **(b)** Lateral view (*top*) shows the extrinsic and intrinsic muscles of the tongue with two tongue floor muscles. Coronal section (*bottom*) shows additional intrinsic muscles

also spreads laterally, reaching a wide area of the tongue root. This bundle draws the tongue root forward and elevates the upper surface of the tongue for high vowels and anterior types of oral consonants. The hyoglossus is a bilateral thin-sheet muscle, which arises from the hyoid bone, runs upward along the sides of the tongue, and ends in the tongue tissue, intermingling with the styloglossus. Its contraction lowers the tongue dorsum and pushes the tongue root backward for the production of low vowels. The styloglossus is a bilateral long muscle originating from the styloid process on the skull base, running obliquely to enter the back sides of the tongue. Within the tongue, it runs forward to reach the apex of the tongue, while branching downward to the hyoid bone and medially toward the midline. Although the extra-lingual bundle of the styloglossus runs obliquely, it pulls the tongue body straight back at the insertion point because the bundle is surrounded by fatty and muscular tissues. The shortening of the intra-lingual bundle draws the tongue apex backward and causes an upward bunching of the tongue body [2.13]. Each of the extrinsic tongue muscles has two functions: drawing of the relevant attachment point toward the origin, and deforming the tongue tissue in the orthogonal orientation. The resulting deformation of the tongue can be explained by two antagonistic pairs of extrinsic muscles: posterior genioglossus versus styloglossus, and anterior genioglossus versus hyoglossus. The muscle arrangement appears to be suitable for tongue body movements in the vertical and horizontal dimensions.

The intrinsic tongue muscle is a group of muscles that have both their origin and termination within the tongue tissue. They include four bilateral muscles: the superior longitudinal, inferior longitudinal, transverse, and vertical muscles. The superior and inferior longitudinal muscles operate on the tongue blade to produce vertical and horizontal movements of the tongue tip. The transverse and vertical muscles together compress the tongue tissue medially to change the cross-sectional shape of the tongue.

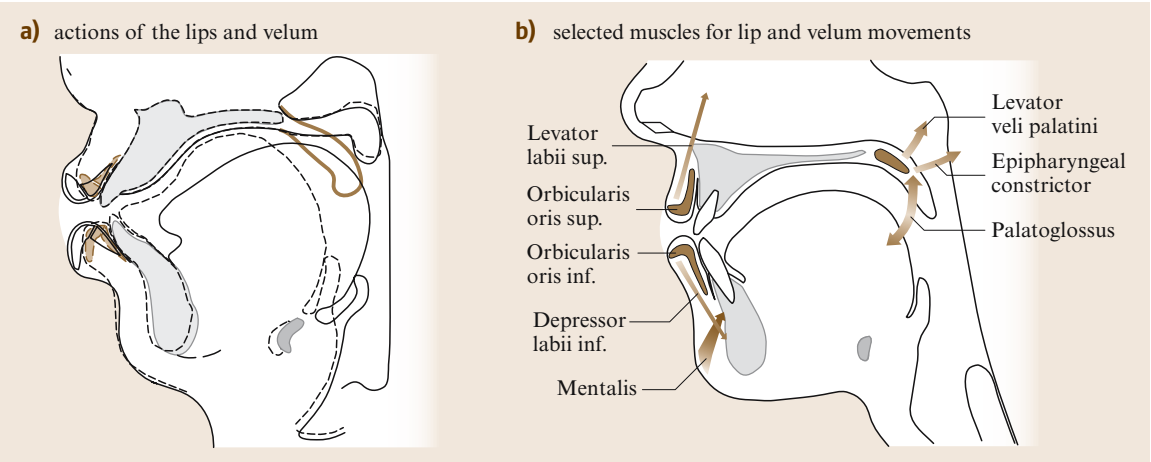

Fig. 2.12a,b Actions of the lips and velum, and their muscles. **(a)** Trace of MRI data in the production of /i/ and /u/ with lip protrusion show that two parts of the orbicularis oris, marginal (*front*) and peripheral (*back*) bundles demonstrate their geometrical changes within the vermillion tissue. The shapes of the velum also vary greatly between the rest position (thick gray line) and vowel articulation. **(b)** Five labial muscles are shown selectively from among many facial muscles. The velum shape is determined by the elevator, constrictor, and depressor (palatopharyngeus)

There are two muscles that support the tongue floor: the geniohyoid and mylohyoid muscles. The geniohyoid runs from the genial process of the mandibular symphysis to the body of the hyoid bone. This muscle has two functions: opening the jaw for open vowels and advancing the hyoid bone to help raise F_0. The mylohyoid is a sheet-like muscle beneath the tongue body that stretches between the mandible and the hyoid bone to support the entire tongue floor. This muscle supports the tongue floor to assist articulation of high front vowels and oral consonants.

Lips and Velum

The lips are a pair of soft-tissue organs consisting of many muscles. Their functions resemble those of the tongue because they partly adhere to the mandible and partly run within the soft tissue of the lips. The vermillion, or the part of red skin, is the unique feature of the human lips, which transmits phonetic signals visually. The deformation of the lips in speech can be divided into three components. The first is opening/closing of the lip aperture, which is augmented by jaw movement. The second is rounding/spreading of the lip tissue, produced by the changes in their left–right dimension. The third is protrusion/retraction of the lip gesture, generated by three-dimensional deformation of the entire lip tissue.

The muscles that cause deformation of the lips are numerous. Figure 2.12 shows only a few representative muscles of the lips. The orbicularis oris is the muscle that surrounds the lips, consisting of two portions; the marginal and peripheral bundles. Contraction of the marginal bundles near the vermillion borders is thought to produce lip rounding without protrusion. Contraction of the peripheral bundles that run in the region around the marginal bundles compresses the lip tissue circumferentially to advance the vermillion in lip protrusion [2.15]. The mentalis arises from the mental part of the mandible to the lip surface, and its contraction elevates the lower lip by pulling the skin at the mental region. The levator labii superior elevates the upper lip, and the depressor labii inferior depresses the lower lip relative to the jaw. The superior and inferior angli oris muscles move the lip corners up and down, respectively, which makes facial expressions rather than speech articulation.

The exact mechanism of lip protrusion is still in question. Tissue bunching by muscle shortening as a general rule for the organs of muscle does not fully apply to the phenomenon of lip protrusion. This is because, as the vermilion thickens in lip protrusion, it does not compress on the teeth; its dental surface often detaches from the teeth (Figure 2.12a). A certain three-dimensional stress distribution within the entire labial tissue must be considered to account for the causal factors of lip protrusion.

The velum, or the soft palate, works as a valve behind the hard palate to control the velopharyngeal port, as shown in Fig. 2.12a. Elevation of the velum is carried

out during the production of oral sounds, while lowering takes place during the production of nasal sounds. The action of the velum to close the velopharyngeal port is not a pure hinge motion but is accompanied by the deformation of the velum tissue with narrowing of the nasopharyngeal wall. In velopharyngeal closure, the levator veli palatine contracts to elevate the velum, and the superior pharyngeal constrictor muscle produces concentric narrowing of the port. In velopharyngeal opening, the palatoglossus muscle assists active lowering of the velum.

2.3.2 Vocal Tract and Nasal Cavity

The vocal tract is an acoustic space where source sounds for speech propagate. Vowels and consonants rely on strengthening or weakening of the spectral components of the source sound by resonance of the air column in the vocal tract. In the broad definition, the vocal tract includes all the air spaces where acoustic pressure variation takes place in speech production. In this sense, the vocal tract divides into three regions: the subglottal tract, the tract from the glottis to the lips, and the nasal cavities.

The subglottal tract is the lower respiratory tract below the glottis down to the lungs via the trachea and bronchial tubes. The length of the trachea from the glottis to the carina is 10–15 cm in adults, including the length of the subglottic laryngeal cavity (about 2 cm). Vocal source sounds propagate from the glottis to the trachea, causing the subglottal resonance in speech spectra. The resonance frequencies of the subglottal airway are estimated to be 640, 1400, and 2100 Hz [2.16]. The second subglottal resonance is often observed below the second formant of high vowels.

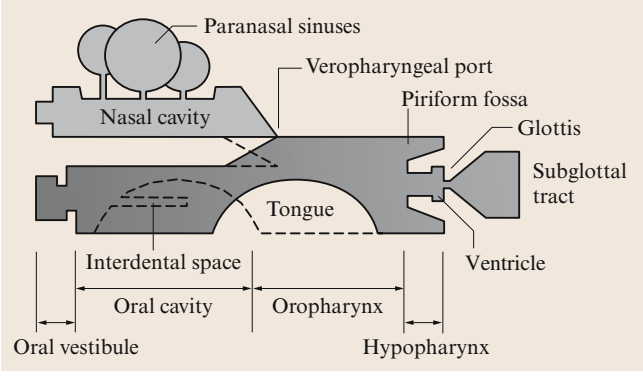

Fig. 2.13 Acoustic design of the vocal tract. Passages from the subglottal tract to two output ends at the lips and nares are shown with the effects of tongue and velar movements. The resonance of the vocal tract above the supraglottic laryngeal cavity determines major the vowel formants (F_1, F_2, and F_3). The resonance of the subglottal tract and interdental space interacts with the vowel formants, while the hypopharyngeal cavities and other small cavities cause local resonances and antiresonances in the higher-frequency region

The vocal tract, according to the conventional definition, is the passage of vocal sounds from the glottis to the lips, where source sounds propagate and give rise to the major resonances. The representative values for the length of the main vocal tract from the glottis to the lips are 15 cm in adult females and 17.5 cm in adult males. According to the measurement data based on the younger population, vocal tract lengths are 14 cm in females and 16.5 cm in males [2.17, 18], which are shorter than the above values. Considering the elongation of the vocal tract during a course of life, the above representative values appear reasonable. While the oral cavity length is maintained by the rigid structures of the skull and jaw, the pharyngeal cavity length increases due to larynx lowering before and after puberty. Thus, elongation of the pharyngeal cavity is the major factor in the developmental variation in vocal tract length.

The vocal tract anatomically divides into four segments: the hypopharyngeal cavities, the mesopharynx, the oral cavity, and the oral vestibule (lip tube). The hypopharyngeal part of the vocal tract consists of the supraglottic laryngeal cavity (2 cm long) and the bilateral conical cavities of the piriform fossa (2 cm long). The mesopharynx extends from the aryepiglottic fold to the anterior palatal arch. The oral cavity is the segment from the anterior palatal arch to the incisors. The oral vestibule extends from the incisors to the lip opening. The latter shows an anterior convexity, which often makes it difficult to measure the exact location of lip opening.

The vocal tract is not a simple uniaxial tube but has a complex three-dimensional construction. The immobile wall of the vocal tract includes the dental arch and the palatal dome. The posterior pharyngeal wall is almost rigid, but it allows subtle changes in convexity and orientation. The soft walls include the entire tongue surface, the velum with the uvula, the lateral pharyngeal wall, and the lip tube. The shape of the vocal tract varies individually due to a few factors. First, the lateral width of the upper and lower jaws relative to the pharyngeal cavity width affects tongue articulation and results in a large individual variation of vocal tract shape observed midsagittally. Second, the mobility of the jaw

depending on the location of the mandibular symphysis relative to the skull can vary the openness of vowels. Third, the size of the tongue relative to the oral and pharyngeal cavities varies individually; the larger the tongue size, the smaller the articulatory space for vowels.

Figure 2.13 shows a schematic drawing of the vocal tract and nasal cavity. The vocal tract has nearly constant branches such as the piriform fossa (entrance to the esophagus) and the vallecula (between the tongue root and epiglottis). The vocal tract also has controlled branches to the nasal cavity at the velopharyngeal port and to the *interdental space* (the space bounded by the upper and lower teeth and the lateral cheek wall). The latter forms a pair of side-branches when the tongue is in a higher position as in /i/ or /e/, while it is unified with the oral cavity when the tongue is in a lower position as in /a/.

The nasal cavity is an accessory channel to the main vocal tract. Its horizontal dimension from the anterior nares to the posterior wall of the epipharynx is approximately 10–11 cm. The nasal cavity can be divided into the single-tube segment (the velopharyngeal region and epipharynx) and the dual-tube segment (the nasal cavity proper and nasal vestibule). Each of the bilateral channels of the nasal cavity proper has a complex shape of walls with the three turbinates with thick mucous membrane, which makes a narrower cross section compared with the epipharyngeal area [2.19]. The nasal cavity has its own side-branches of the paranasal sinuses; the maxillary, sphenoid, ethomoid, and frontal sinuses.

The nasal cavity builds nasal resonance to accomplish phonetic features of nasal sounds and nasalized vowels. The paranasal sinuses also contribute to acoustic characteristics of the nasal sounds. The nasal murmur results from these characteristics: a Helmholtz resonance of the entire nasopharyngeal tract from the glottis to the anterior nares and regional Helmholtz resonances caused by the paranasal sinuses, together characterized by a resonance peak at 200–300 Hz and spectral flattening up to 2 kHz [2.20, 21]. The nasal resonance could takes place even in oral vowels with a complete closure of the velopharyngeal port: the soft tissue of the velum transmits the pressure variation in the oral cavity to the nasal cavity, which would enhances sound radiation for close vowels and voiced stops.

2.3.3 Aspects of Articulation in Relation to Voicing

Here we consider a few phonetic evidences that can be considered as joint products of articulation and phonation. Vowel production is the typical example for this topic, in view of its interaction with the larynx. Regulation of voice quality, which has been thought to be a laryngeal phenomenon, is largely affected by the lower part of the vocal tract. The voiced versus voiceless distinction is a pertinent issue of phonetics that involves both phonatory and articulatory mechanisms.

Production of Vowels

The production of vowels is the result of the joint action of phonatory and articulatory mechanisms. In this process, the larynx functions as a source generator, and the vocal tract plays the role of an acoustic filter to modulate the source sounds and radiate from the lip opening, as described by the *source-filter* theory [2.22, 23]. The quality of oral vowels is determined by a few peak frequencies of vocal tract resonance (formants). In vowel production, the vocal tract forms a *closed tube* with the closed end at the glottis and the open end at the lip opening. Multiple reflections of sound wave between the two ends of the vocal tract give rise to vowel formants (F_1, F_2, F_3). The source-filter theory has been supported by many studies as the fundamental concept explaining the acoustic process of speech production, which is further discussed in the next section.

Vowel articulation is the setup for the articulatory organs to determine vocal tract shape for each vowel. When the jaw is in a high position and the tongue is in a high front position, the vocal tract assumes the shape for /i/. Contrarily, when the jaw is in a low position and the tongue is in a low back position, the vocal tract takes the shape for /a/. The articulatory organ that greatly influences vocal tract shape for vowels is the tongue. When the vocal tract is modeled as a tube with two segments (front and back cavities), the movements of the tongue body between its low back and high front positions creates contrasting diverging and converging shapes of the main vocal tract. Jaw movement enhances these changes in the front cavity volume, while pharyngeal constriction assists in the back cavity volume. When the vocal tract is modeled as a tube with three segments, the movements of the tongue body between its high back and low front positions determine the constriction or widening of the vocal tract in its middle portion. The velum also contributes to the articulation of open vowels by decreasing the area of the vocal tract at the velum or making a narrow branch to the nasal cavity. The lip tube is another factor for vowel articulation that determines the vocal tract area near the open end.

Although muscular control for vowel articulation is complex, a simplified view can be drawn based

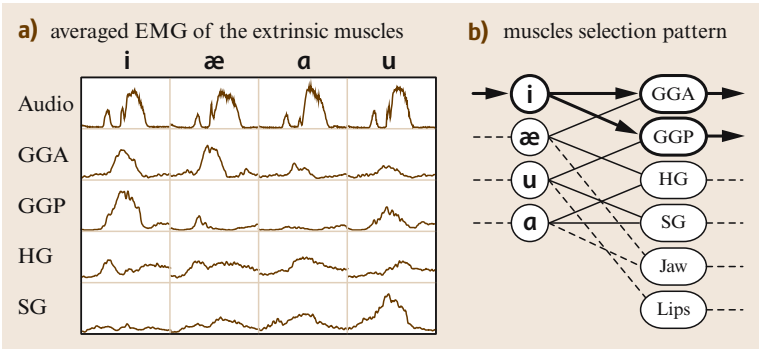

Fig. 2.14a,b Tongue EMG data during VCV utterances and muscle selection pattern in vowel articulation. **(a)** Averaged EMG data for four English corner vowels are shown for the major muscles of the tongue: the anterior genioglossus (GGA), posterior genioglossus (GGP), hyoglossus (HG), and styloglossus (SG). **(b)** The systematic variation observed in the muscle–vowel matrix suggests a muscle selection pattern

on electromyographic (EMG) data obtained from the tongue muscles [2.24]. Figure 2.14a shows a systematic pattern of muscle activities for CVC (consonant-vowel-consonant) utterances with /p/ and four English corner vowels. The anterior and posterior genioglossus are active for front vowels, while the styloglossus and hyoglossus are active for back vowels. These muscles also show a variation depending on vowel height. These observations are shown schematically in Fig. 2.14b: the basic control pattern for vowel articulation is the selection of two muscles among the four extrinsic muscles of the tongue [2.25].

As the tongue or jaw moves for vowel articulation, they apply forces to the surrounding organs and cause secondary effects on vowel sounds. For example, articulation of high vowels such as /i/ and /u/ is mainly produced by contraction of the posterior genioglossus, which is accompanied by forward movement of the hyoid bone. This action applies a force to rotate the thyroid cartilage in a direction that stretches the vocal folds. In evidence, higher vowels tend to have a higher F_0, known as the *intrinsic vowel* F_0 [2.26,27]. When the jaw opens to produce open vowels, jaw rotation compresses the tissue behind the mandibular symphysis, which applies a force to rotate the thyroid cartilage in the opposite direction, thereby shortening the vocal folds. Thus, the jaw opening has the secondary effect of lowering the intrinsic F_0 for lower vowels.

Supra-Laryngeal Control of Voice Quality

The laryngeal mechanisms controlling voice quality were described in an earlier section. In this section, the supra-laryngeal factors are discussed. Recent studies have shown evidence that the resonances of the hypopharyngeal cavities determine the spectral envelope in the higher frequencies above 2.5 kHz by causing an extra resonance and antiresonances [2.28–31]. The hypopharyngeal cavities include a pair of vocal-tract

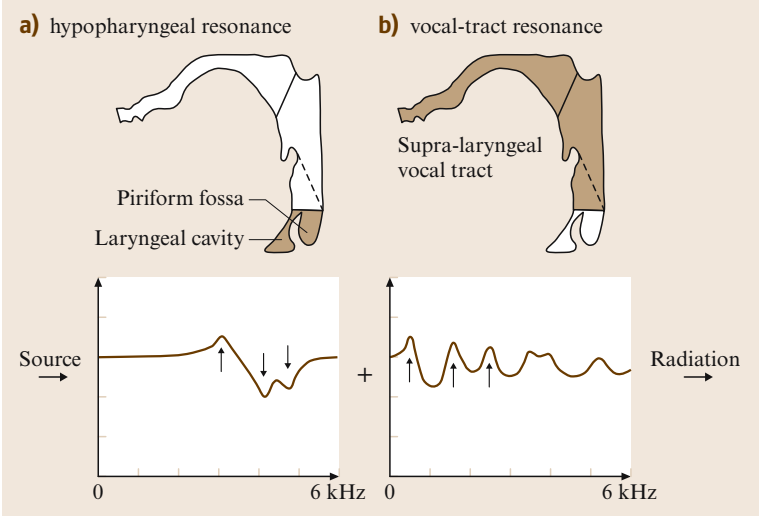

Fig. 2.15a,b Vocal-tract resonance with hypopharyngeal cavity coupling in vowel production. **(a)** The supraglottal laryngeal cavity contributes a resonance peak at 3–3.5 kHz, and the bilateral cavities of the piriform fossa cause antiresonances at 4–5 kHz. **(b)** The main vocal tract above the laryngeal cavity determines the major vowel formants

side-branches formed by the piriform fossa. Each fossa maintains a relatively constant cavity during speech, which is collapsed only in deep inhalation by the wide abduction of the arytenoid cartilage. The piriform fossa causes one or two obvious antiresonances in the higher frequencies above 4 kHz [2.29] and affects the surrounding formants. The laryngeal cavity above the vocal folds also contributes to shaping the higher frequencies [2.28, 32]. The supraglottic laryngeal cavity, from the ventricles to the aryepiglottic folds via the ventricular folds, forms a type of Helmholtz resonator and gives rise to a resonance at higher frequencies of 3–3.5 kHz. This resonance can be counted as the fourth formant (F4) but it is actually an *extra formant* to the resonance of the vocal tract above the laryngeal cavity [2.30]. When the glottis opens in the open phase of vocal fold vibration, the supraglottic laryngeal cavity no longer constitutes a typical Helmholtz resonator, and demonstrates a strong damping of the resonance, which is observed as the disappearance of the affiliated extra formant. Therefore, the laryngeal cavity resonance shows a cyclic nature during vocal fold vibration, and it is possibly absent in breathy phonation or pathological conditions with insufficient glottal closure [2.31]. Figure 2.15 shows an acoustic model of the vocal tract to illustrate this coupling of the hypopharyngeal cavities.

The hypopharyngeal cavities are not an entirely fixed structure but vary due to physiological efforts to control F_0 and voice quality. A typical case of the hypopharyngeal adjustment of voice quality is found in the *singing formant* [2.28]. When high notes are produced by opera singers, the entire larynx is pulled forward due to the advanced position of the tongue, which widens the piriform fossa to deepens the fossa's antiresonances, resulting in a decrease of the frequency of the adjacent lower formant (F_5). When the supraglottic laryngeal cavity is constricted, its resonance (F_4) comes down towards the lower formant (F_3). Consequently, the third to fifth formants come closer to each other and generate a high resonance peak observed near 3 kHz.

Regulation of Voiced and Voiceless Sounds

Voiced and voiceless sounds are often attributed to the glottal state with and without vocal fold vibration, while their phonetic characteristics result from phonatory and articulatory controls over the speech production system. In voiced vowels, the vocal tract forms a closed tube with no significant constrictions except for the narrow laryngeal cavity. On the other hand, in whispered vowels, the membranous glottis is closed, and the supraglottic laryngeal cavity forms an extremely narrow channel continued from the open cartilaginous glottis, with a moderate constriction of the lower pharynx. Devoiced vowels exhibit a wide open glottis and a reduction of tongue articulation. Phonetic distinctions of voiced and voiceless consonants further involve fine temporal control over the larynx

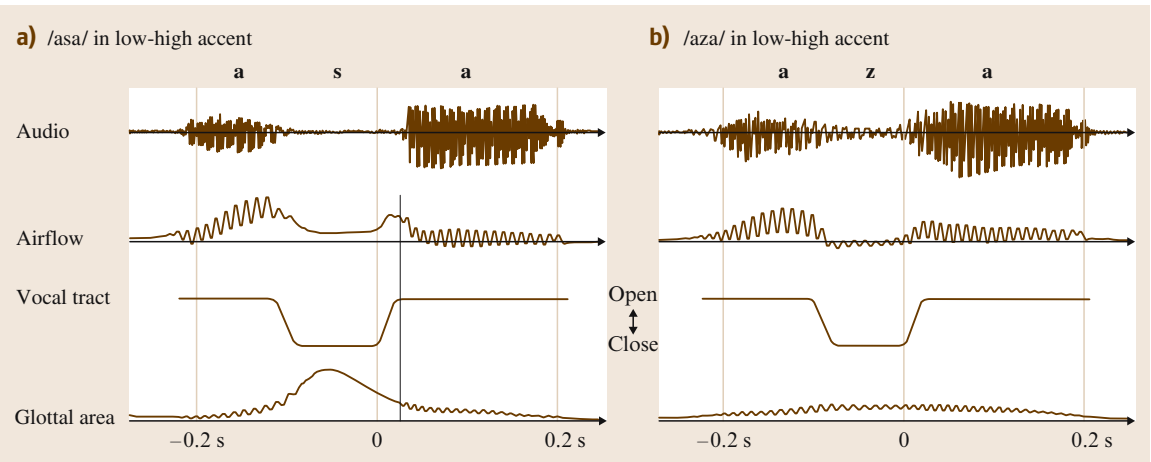

Fig. 2.16a,b Laryngeal articulatory patterns in producing VCV utterances with voiceless and voiced fricatives as in /asa/ and /aza/. From the top to bottom, speech signals, oral airflow, schematic patterns of vocal tract constriction, and glottal area variations are shown schematically. This figure is based on the author's recent experiment with anemometry with an open-type airflow transducer and photoglotto-graphy with an external lighting technique, conducted by *Dr. Shinji Maeda* (ENST) and the author

and supra-laryngeal articulators in language-specific ways.

In the production of voiced consonants, vocal fold vibration typically continues during the voiced segments. In voiced stops and fricatives, the closure or narrowing of the vocal tract results in decrease in glottal airflow and transglottal pressure difference. The glottal airflow during the stop closure is maintained during the closure due to the increases in vocal tract volume: the expansion of the oral cavity (jaw lowering and cheek wall expansion) and the expansion of the pharyngeal cavity (lateral wall expansion and larynx lowering). During the closure period, air pressure variations are radiated not only from the vocal tract wall but also from the anterior nares due to transvelar propagation of the intra-oral sound pressure into the nasal cavities.

In the production of voiceless consonants, vocal fold vibration is suppressed due to a rapid reduction of the transglottal pressure difference and abduction of the vocal folds. During stop closures, the intra-oral pressure builds up to reach the subglottal pressure, which enhances the rapid airflow after the release of the closure. Then, vocal fold vibration restarts with a delay to the release, which is observed as a long voice onset time (VOT) for voiceless stops. The process of suppressing vocal fold vibration is not merely a passive aerodynamic process on the vocal folds, but is assisted by a physiological process to control vocal fold stiffness. The cricothyroid muscle has been observed to increase its activity in producing voiceless consonants. This activity results in a high–falling F_0 pattern during the following vowel, contributing a phonetic attribute to voiceless consonants [2.33]. In glottal stops, vocal fold vibration stops due to forced adduction of the vocal folds with an effort closure of the supraglottic laryngeal cavity.

Figure 2.16 illustrates the time course of the processes during vowel-consonant-vowel (VCV) utterances with a voiceless fricative in comparison to the case with a voiced fricative. The voiceless segment initiates with glottal abduction and alveolar constriction, and vocal fold vibration gradually fades out during the phase of glottal opening. After reaching the maximum glottal abduction, the glottis enters the adduction phase, followed by the release of the alveolar constriction. Then, the glottis becomes narrower and vocal fold vibration restarts. There is the time lag between the release of the constriction and full adduction of the glottis, which results in the peak flow seen in Fig. 2.16a, presumably accompanied by aspiration sound at the glottis.

2.3.4 Articulators' Mobility and Coarticulation

The mobility of speech articulators varies across organs and contributes certain phonetic characteristics to speech sounds. Rapid movements are essential to a sequence from one distinctive feature to another, as observed in the syllable /sa/ from a narrow constriction to the vocalic opening, while gradual movements are found to produce nasals and certain labial sounds. These variations are principally due to the nature of articulators with respect to their mobility. The articulatory mechanism involves a complex system that is built up by organs with different motor characteristics. Their variation in temporal mobility may be explained by a few biological factors. The first is the phylogenetic origin of the organs: the tongue muscles share their origin with the fast motor systems such as the eyeball or finger, while other muscles such as in the lips or velum originate from the slow motor system similar to the musculature of the alimentary tract. The second is the innervation density to each muscle: the faster organs are innervated by thicker nerve bundles, and vise versa, which derives from an adaptation of the biological system to required functions. In fact, the human hypoglossal nerve that supplies the tongue muscles is much thicker than that of other members of the primate family. The third is the composition of muscle fiber types in the musculature, which varies from organ to organ. The muscles in the larynx have a high concentration of the ultrafast fibers (type 2B), while the muscle to elevate the velum predominantly contains the slow fibers (type 1). In accordance with these biological views, the rate of the articulators movement indexed by the maximum number of syllables per second follows the order of the tongue apex, body, and lips: the tongue moves at a maximum rate of 8.2 syllables per second at the apex, and 7.1 syllables per second with the back of the tongue, while the lips and facial structures move at a maximum rate of 2.5–3 syllables per second [2.34]. More recent measurements indicate that the lips are slower than the tongue apex but faster than the tongue dorsum. The velocities during speech tasks reach 166 mm/sec for the lower lip, 196 mm/sec for the tongue tip, and 129 mm/sec for the tongue dorsum [2.35]. The discrepancy between these two reports regarding the mobility of the lips may be explained by the types of movements measured: opening–closure movement by the jaw–lower lip complex is faster than the movement of the lips themselves, such as protrusion and spreading.

It is often noted that speech is characterized by asynchrony among articulatory movements, and the degree

of asynchrony varies with the feature to be realized. Each articulator does not necessarily strictly keep pace with other articulators in a syllable sequence. The physiological basis of this asynchrony may be explained by the mobility of the articulatory organs and motor precision required for the target of articulation. The slower articulators such as the lips and velum tend to exhibit marked coarticulation in production of labial and nasal sounds. In stop–vowel–nasal sequences (such as /tan/), the velopharyngeal port is tightly closed at the stop onset and the velum begins to lower before the nasal consonant. Thus, the vowel before the nasal consonant is partly nasalized. When the vowel /u/ is preceded by /s/, the lips start to protrude during the consonant prior to the rounded vowel.

The articulators' mobility also contributes some variability to speech movements. The faster articulators such as parts of the tongue show various patterns from target undershooting to overshooting. In articulation of close–open–close vowel sequences such as /iai/, tongue movements naturally show undershooting for the open vowel. In contrast, when the alveolar voiceless stop /t/ is placed in the open vowel context as in /ata/, the tongue blade sometimes shows an extreme overshoot with a wide contact on the hard palate because such articulatory variations do not significantly affect the output sounds. On the contrary, in alveolar and postalveolar fricatives such as /s/ and /sh/, tongue movements also show a dependence on articulatory precision because the position of the tongue blade must be controlled precisely to realize the narrow passage for generating frication sounds. The lateral /l/ is similar to the stops with respect to the palatal contact, while the rhotic /r/ with no contact to the palate can show a greater extent of articulatory variations from retroflex to bunched types depending on the preceding sounds.

2.3.5 Instruments for Observing Articulatory Dynamics

X-ray and palatography have been used as common tools for articulatory observation. Custom instruments are also developed to monitor articulatory movements, such as the X-ray microbeam system and magnetic sensor system. The various types of newer medical imaging techniques are being used to visualize the movements of articulatory system using sonography and nuclear magnetic resonance. These instruments are generally large scale, although relatively compact instruments are becoming available (e.g., magnetic probing system or portable ultrasound scanner).

Palatography

The palatograph is a compact device to record temporal changes in the contact pattern of the tongue on the palate. There are traditional static and modern dynamic types. The dynamic type is called electropalatography, or dynamic palatography, which employs an individually customized palatal plate to be placed on the upper jaw. As shown in Fig. 2.17a, this system employs a palatal plate with many surface electrodes to monitor electrical contacts on the tongue's surface.

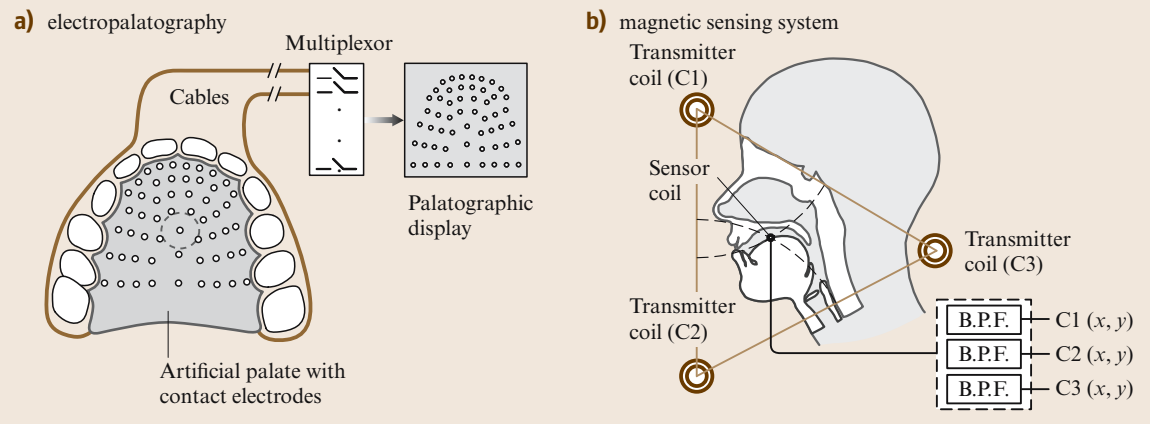

Fig. 2.17a,b Electropalatography and magnetic sensing system. **(a)** Electropalatography displays tongue–palate contact patterns by detecting weak electrical current caused by the contact between the electrodes on the artificial palate and the tongue tissue. **(b)** Magnetic sensing system is based on detection of alternate magnetic fields with different frequencies using miniature sensor coils

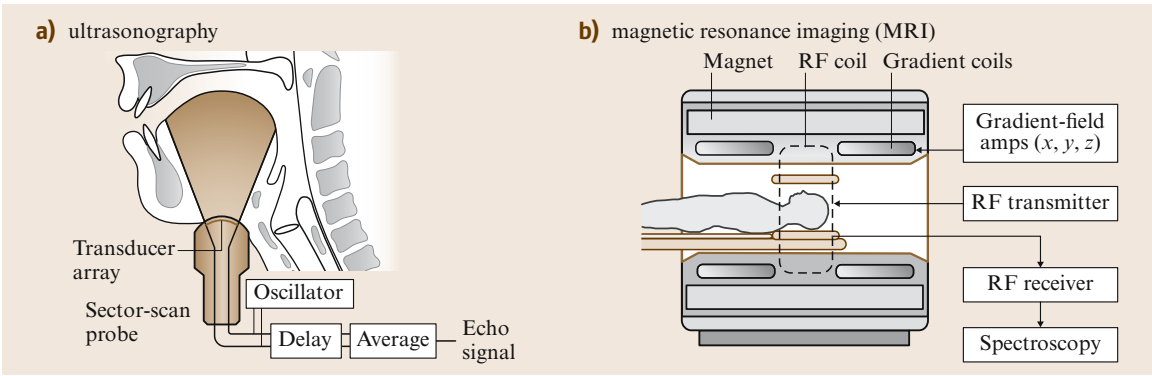

Fig. 2.18a,b Medical imaging techniques. **(a)** Ultrasound scanner uses an array of transmitters and receivers to detect echo signals from regions where the ultrasound signals reflect strongly such as at the tissue-air boundaries on the tongue surface. **(b)** Magnetic resonance imaging (MRI) generates strong static magnetic field, controlled gradient fields in the three directions, and radio-frequency (RF) pulses. Hydrogen atoms respond to the RF pulses to generate echo signals, which are detected with a receiver coil for spectral analysis

Marker Tracking System

A few custom devices have been developed to record movements of markers attached on the articulatory organs. X-ray microbeam and magnetic sensing systems belong to this category. Both can measure 10 markers simultaneously. The X-ray microbeam system uses a computer-controlled narrow beam of high-energy X-rays to track small metal pellets attached on the articulatory organs. This system allows automatic accurate measurements of pellets with a minimum X-ray dosage.

The magnetic sensing system (magnetometer, or magnetic articulograph) is designed to perform the same function as the microbeam system without X-rays. The system uses a set of transmitter coils that generate alternate magnetic fields and miniature sensor coils attached to the articulatory organs, as shown in Fig. 2.17b. The positions of the receiver coils are computed from the filtered signals from the coils.

Medical Imaging Techniques

X-ray cinematography and X-ray video fluorography have been used for re-cording articulatory movements in two-dimensional projection images. The X-ray images show clear outlines of rigid structures, while they provide less-obvious boundaries for soft tissue. The outline of the tongue is enhanced by the application of liquid contrast media on the surface. Metal markers are often used to track the movements of flesh points on the soft-tissue articulators.

Ultrasonography is a diagnostic technique to obtain cross-sectional images of soft-tissues in real time. Ultrasound scanners consist of a sound probe (phased-array piezo transducer and receiver) and image processor, as illustrated in Fig. 2.18a. The probe is attached to the skin below the tongue to image the tongue surface in the sagittal or coronal plane.

Magnetic resonance imaging (MRI), shown in Fig. 2.18b, is a developing medical technique that excels at soft-tissue imaging of the living body. Its principle relies on excitation and relaxation of the hydrogen nuclei in water in a strong homogeneous magnetic field in response to radio-frequency (RF) pulses applied with variable gradient magnetic fields that determine the slice position. MRI is essentially a method for recording static images, while motion imaging setups with stroboscopic or real-time techniques have been applied to the visualization of articulatory movements or vocal tract deformation three-dimensionally [2.36].

2.4 Summary

This chapter described the structures of the human speech organs and physiological mechanisms for producing speech sounds. Physiological processes during speech are multidimensional in nature as described in this chapter. Discoveries of their component mechanisms have been dependent on technical developments

for visualizing the human body and analyses of biological signals, and this is still true today. For example, the hypopharyngeal cavities have long been known to exist, but their acoustic role was underestimated until recent MRI observations. The topics in this chapter were chosen with the author's hope to provide a guideline for the sophistication of speech technologies by reflecting the real and detailed processes of human speech production. Expectations from these lines of studies include speech analysis by recovering control parameters of articulatory models from speech sounds, speech synthesis with full handling of voice quality and individual vocal characteristics, and true speech recognition through biologic, acoustic, and phonetic characterizations of input sounds.

References

2.1 M.H. Draper, P. Ladefoged, D. Whittenridge: Respiratory muscles in speech, J. Speech Hearing Res. **2**, 16–27 (1959)

2.2 T.J. Hixon, M. Goldman, J. Mead: Kinematics of the chest wall during speech production: volume displacements of the rib cage, abdomen, and lung, J. Speech Hearing Res. **16**, 78–115 (1973)

2.3 G. Weismer: Speech production. In: *Handbook of Speech-Language Pathology and Audiology*, ed. by N.J. Lass, L.V. McReynolds, D.E. Yoder (Decker, Toronto 1988) pp. 215–252

2.4 J. Kahane: A morphological study of the human prepubertal and pubertal larynx, Am. J. Anat. **151**, 11–20 (1979)

2.5 M. Hirano, Y. Kakita: Cover-body theory of vocal cord vibration. In: *Speech Science*, ed. by R. Daniloff (College Hill, San Diego 1985) pp. 1–46

2.6 E.B. Holmberg: Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice, J. Acoust. Soc. Am. **84**, 511–529 (1988)

2.7 M.R. Rothenberg: Acoustic interaction between the glottal source and the vocal tract. In: *Vocal Fold Physiology*, ed. by K.N. Stevens, M. Hirano (Univ. Tokyo Press, Tokyo 1981) pp. 305–328

2.8 G. Fant, J. Liljencrants, Q. Lin: A four-parameter model of glottal flow, Speech Transmission Laboratory – Quarterly Progress and Status Report (STL-QPSR) **4**, 1–13 (1985)

2.9 B.R. Fink, R.J. Demarest: *Laryngeal Biomechanics* (Harvard Univ. Press, Cambridge 1978)

2.10 J.E. Atkinson: Correlation analysis of the physiological features controlling fundamental frequency, J. Acoust. Soc. Am. **63**, 211–222 (1978)

2.11 K. Honda, H. Hirai, S. Masaki, Y. Shimada: Role of vertical larynx movement and cervical lordosis in F0 control, Language Speech **42**, 401–411 (1999)

2.12 H. Takemoto: Morphological analysis of the human tongue musculature for three-dimensional modeling, J. Speech Lang. Hearing Res. **44**, 95–107 (2001)

2.13 S. Takano, K. Honda: An MRI analysis of the extrinsic tongue muscles during vowel production, Speech Commun. **49**, 49–58 (2007)

2.14 S. Maeda: Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In: *Speech Production and Speech Modeling*, ed. by W.J. Hardcastle, A. Marchal (Kluwer Academic, Dartrecht 1990) pp. 131–149

2.15 K. Honda, T. Kurita, Y. Kakita, S. Maeda: Physiology of the lips and modeling of lip gestures, J. Phonetics **23**, 243–254 (1995)

2.16 K. Ishizaka, M. Matsudaira, T. Kaneko: Input acoustic-impedance measurement of the subglottal system, J. Acoust. Soc. Am. **60**, 190–197 (1976)

2.17 U.G. Goldstein: An articulatory model for the vocal tracts of growing children. Ph.D. Thesis (Massachusetts Institute of Technology, Cambridge 1980)

2.18 W.T. Fitch, J. Giedd: Morphology and development of the human vocal tract: A study using magnetic resonance imaging, J. Acoust. Soc. Am. **106**, 1511–1522 (1999)

2.19 J. Dang, K. Honda, H. Suzuki: Morphological and acoustic analysis of the nasal and paranasal cavities, J. Acoust. Soc. Am. **96**, 2088–2100 (1994)

2.20 O. Fujimura, J. Lindqvist: Sweep-tone measurements of the vocal tract characteristics, J. Acoust. Soc. Am. **49**, 541–557 (1971)

2.21 S. Maeda: The role of the sinus cavities in the production of nasal vowels, Proc. IEEE Int. Conf. Acoustics, Speech, Signal Proc. (ICASSP'82), Vol. 2 (1982) pp. 911–914, Paris

2.22 T. Chiba, M. Kajiyama: *The Vowel – Its Nature and Structure* (Tokyo-Kaiseikan, Tokyo 1942)

2.23 G. Fant: *Acoustic Theory of Speech Production* (Mouton, The Hague 1960)

2.24 T. Baer, P. Alfonso, K. Honda: Electromyography of the tongue muscle during vowels in /*pVp/ environment, Ann. Bull. RILP **22**, 7–20 (1988)

2.25 K. Honda: Organization of tongue articulation for vowels, J. Phonetics **24**, 39–52 (1996)

2.26 I. Lehiste, G.E. Peterson: Some basic considerations in the analysis of intonation, J. Acoust. Soc. Am. **33**, 419–425 (1961)

2.27 K. Honda: Relationship between pitch control and vowel articulation. In: *Vocal Fold Physiology*, ed. by D.M. Bless, J.H. Abbs (College-Hill, San Diego 1983) pp. 286–297

2.28 J. Sundberg: Articulatory interpretation of the singing formant, J. Acoust. Soc. Am. **55**, 838–844 (1974)

2.29 J. Dang, K. Honda: Acoustic characteristics of the piriform fossa in models and humans, J. Acoust. Soc. Am. **101**, 456–465 (1996)
2.30 H. Takemoto, S. Adachi, T. Kitamura, P. Mokhtari, K. Honda: Acoustic roles of the laryngeal cavity in vocal tract resonance, J. Acoust. Soc. Am. **120**, 2228–2238 (2006)
2.31 T. Kitamura, H. Takemoto, S. Adachi, P. Mokhtari, K. Honda: Cyclicity of laryngeal cavity resonance due to vocal fold vibration, J. Acoust. Soc. Am. **120**, 2239–2249 (2006)
2.32 I.R. Titze, B.H. Story: Acoustic interactions of the voice source with the lower vocal tract, J. Acoust. Soc. Am. **101**, 2234–2243 (1997)
2.33 A. Lofqvist, N.S. McGarr, K. Honda: Laryngeal muscles and articulatory control, J. Acoust. Soc. Am. **76**, 951–954 (1984)
2.34 R.G. Daniloff: Normal articulation processes. In: *Normal Aspect of Speech, Hearing, and Language*, ed. by F.D. Minifie, T.J. Hixon, F. Williams (Prentice-Hall, Englewood Cliffs 1983) pp. 169–209
2.35 D.P. Kuehn, K.L. Moll: A cineradiographic study of VC and CV articulatory velocities, J. Phonetics **4**, 303–320 (1976)
2.36 K. Honda, H. Takemoto, T. Kitamura, S. Fujita, S. Takano: Exploring human speech production mechanisms by MRI, IEICE Info. Syst. **E87-D**, 1050–1058 (2004)