Chapter 95: Physiology

Colin Painter

Mechanisms of Normal and Abnormal Swallowing

Jeri A. Logemann

Stages of Normal Deglutition

Swallowing is a complex neuromuscular function involving structures in the oral cavity, pharynx, larynx, and esophagus. Various researchers have divided normal swallowing into four stages: the oral preparatory stage, the oral stage, the pharyngeal stage, and the esophageal stage (Ardran and Kemp, 1951; Bosma, 1957; Fyke and Code, 1955; Logemann, 1983; Mandelstam and Liever, 1970). The first two stages, the oral preparatory and the oral, are under voluntary control, whereas the second two stages, the pharyngeal and esophageal, are involuntary, being under reflexive control (Miller, 1972; Sessle and Hannan, 1976).

Voluntary stages

Oral preparatory stages

The oral preparatory aspect of deglutition involves the coordination of (1) lip closure to hold food in the mouth anteriorly, (2) tension in the labial and buccal musculature to close the anterior and lateral sulci, (3) rotary motion of the jaw for chewing, (4) lateral rolling motion of the tongue to position food on the teeth during mastication, and (5) bulging forward of the soft palate to seal the oral cavity posteriorly and widen the nasal airway, which prepares food for the swallow (Logemann, 1983). Much of the coordination requires cerebellar input (Larson and Sutton, 1978). This phase of deglutition is mechanical in that it involves the reduction of solid food to a pulverized consistency that can easily be swallowed. The most important neuromuscular function involved in the oral preparatory phase is the lateral rolling motion of the tongue (Ingervall, 1978). Without normal range of tongue motion, the manipulation and mastication of food during the oral preparatory phase would be impossible. Although patients can chew without teeth and with restricted jaw motion, reduced buccal tension, or reduced lip closure, they cannot chew without normal tongue mobility. At the termination of the oral preparatory phase, the tongue pulls the food together into a ball or bolus and holds it in a cohesive fashion against the hard palate (Fig. 95-1), in preparation for the beginning of the oral phase of the swallow (Logemann, 1983; Shawker et al, 1983; Shedd et al, 1961).

Oral stage

The oral stage of the swallow is also a mechanical stage and is designed to move food from the front of the oral cavity to the anterior faucial arches, where the reflexive swallow is initiated. Again, tongue motion is the most critical element in this phase of the swallow, since the tongue shapes, lifts, and squeezes the bolus upward an backward along the hard palate until the food reaches the anterior faucial arches (Hrycyshyn and Basmajian, 1972; Lower, 1981). Tension in the buccal musculature is also thought to contribute to propelling the bolus backward, but to a much lesser degree than tongue movements (Shedd et al, 1960). At the point of the anterior faucial arches the swallowing reflex is triggered in the normal individual (Fig. 95-2). When this reflex triggers the glossopharyngeal nerve (C9), the oral phase of the swallow is terminated and the pharyngeal or reflexive stage is initiated. In normal persons the oral stage lasts no longer than 1 second and does not vary significantly with age, sex, or consistency of the bolus swallowed (Blonsky et al, 1975; Mandelstam and Lieber, 1970).

In animal studies the swallowing reflex has been found to be triggered not only from the glossopharyngeal nerve (C9) but also from the superior laryngeal nerve at the inlet of the larynx. The normal human does not use this second mechanism for triggering the reflex, since food would have almost entered the airway by the time the reflex was triggered. Thus when videofluoroscopic examination of a patient shows that the reflex triggers late, as food is about the enter the airway or after the bolus has come to rest in the valleculae or piriform sinuses, the patient is said to have a disorder of swallowing called a delayed swallowing reflex.

Involuntary stages

The reflexive aspect of the swallow is the more important because airway protection is maintained during this stage in the normal individual. The oral preparatory and oral stages of the swallow can be bypassed by reducing the consistency of food to liquid, by syringing food into the back of the mouth, or by positioning the head back so that gravity carries the food into the pharynx. The pharyngeal or reflexive stage of the swallow cannot be bypassed.

The swallowing reflex is mediated in the brainstem in the reticular formation immediately adjacent to the respiratory center. Coordination exists between these two centers, since respiration ceases for a fraction of a second during the reflexive swallow. There is also cortical input to the triggering of the swallowing reflex through the tongue movement patterns in the oral phase of the swallow (Jean and Car, 1979). The exact neurological substrate for this cortical input is not clearly understood. However, the importance of this input is illustrated in patients who suffer cerebrovascular accidents localized to the cortex. The vast majority of these patients who exhibit swallowing disorders have a delayed triggering or absence of the swallowing reflex (Veis and Logemann, 1983).

Neuromuscular activities

When the swallowing reflex is triggered, the brainstem swallowing center programs four neuromuscular activities to occur (Fig. 95-3). If the central processing mechanism in the brainstem is damage, as is often seen in neurologically impaired patients such as those who have suffered a brainstem stroke, the reflex is not triggered and none of the four neuromuscular functions occurs. Some patients with head and neck cancer have damage to the sensory input to the brainstem center, which results in delayed or absent triggering of the reflex. If triggering of the reflex is delayed, none of these neuromuscular activities will occur until the reflex triggers (1) velopharyngeal closure to prevent reflux of material up the nose; (2) pharyngeal peristalsis to propel the bolus through the pharynx; (3) airway protection, which involves both *elevation* and *closure* of the larynx; and (4) cricopharyngeal or upper esophageal sphincter opening to allow the bolus to pass into the esophagus (Ardran and Kemp, 1951; Bosma, 1957; Negus, 1949).

Although these four functions are often described as occurring almost simultaneously, preliminary data from the frame-by-frame analysis of videofluoroscopic studies of swallowing indicate that they occur in a rapid sequence from 1 to 4. That is, velopharyngeal closure is complete as the first response to triggering the reflex, in conjunction with the onset of pharyngeal peristalsis. Then laryngeal closure occurs, followed by relaxation of the cricopharyngeal sphincter. These neuromuscular functions also overlap but do not all last for the entire pharyngeal stage of the swallow. Instead each lasts only as long as the bolus is passing that part of the pharynx. For example, the soft palate remains closed only as long as the bolus is passing the velopharyngeal opening. When the bolus has reached the middle to lower pharynx, the soft palate lowers. At that time the larynx is closed and elevated, preventing the passing bolus from entering the airway. The duration of the pharyngeal stage of the swallow (beginning when the swallowing reflex triggers at the anterior faucial arches and terminating when the bolus passes through the cricopharyngeus muscle) is a maximum of 1 second and does not vary significantly with consistency of food or age or sex of the individual (Blonsky et al, 1975; Mandelstam and Lieber, 1970).

Airway protection

Several of the neuromuscular activities that are programme by the swallowing reflex deserve further attention. Airway protection, as note previously, involves two dimensions: elevation and closure. The elevation is created by contraction of strap musculature, which positions the larynx upward and forward under the tongue as the tongue base is retracted at the end of the oral phase of the swallow. Thus the larynx is pulled up and out of the way of the passage of the food bolus over the base of the tongue. Closure of the larynx is thought to involve three sphincters: the epiglottis and the aryepiglottic folds, the false vocal folds, and the true vocal folds (Ardran and Kemp, 1952, 1956, 1967). The relative importance of each of these sphincters has been debated in the literature (Ardran and Kemp, 1967; Fink, 1975; Fink and Demarest, 1978). Most authors agree that the epiglottis and aryepiglottic folds play a relatively minor role in protecting the airway and that the true vocal folds play a major role. It is important to note that airway closure is maintained only for the fraction of a second that the bolus is passing the airway and that the airway opens for inhalation after the bolus has entered the esophagus.

Pharyngeal peristalsis

Electromyography (EMG) studies have shown that pharyngeal peristalsis, or the squeezing action of the constrictor mechanism to move the bolus through the pharynx, occurs sequentially, beginning in the superior constrictor muscle and moving through the medial to the inferior constrictor muscle (Doty and Bosma, 1956). The relationship of the contraction of the constrictors to the timing of the opening of the cricopharyngeus region is not clearly understood. Possibly the progressive contraction of these constrictors acts as a timing mechanism to trigger the opening of the cricopharyngeus region.

Pharyngeal peristalsis is responsible for clearing material from the pharyngeal recesses, including the valleculae and the piriform sinuses. When residue or material remains in the valleculae or in both the valleculae and the piriform sinuses after the swallow, it is interpreted as a symptom of reduced peristalsis (Logemann, 1983).

Cricopharyngeal action

The cricopharyngeus muscle acts in opposition to the function of the constrictor mechanism of the pharynx. At rest the constrictors are relaxed and the cricopharyngeus muscle or upper esophageal sphincter is closed to prevent air intake into the esophagus concurrent with inhalation into the lungs. Also, the closed cricopharyngeus region prevents reflux from the esophagus into the pharynx (Ardran and Kemp, 1951). During the swallow, as the constrictor mechanism is contracting, the cricopharyngeal region opens at the appropriate moment to allow the bolus to pass into the esophagus an immediately recloses to prevent reflux. The exact mechanism operating to open the cricopharyngeal region is just beginning to be understood. Factors that have been identified as contributing to this opening are (1) relaxation of the cricopharyngeal muscle; (2) the upward, forward pull of the larynx, which acts to open the sphincter; and (3) bolus pressure to widen the opening.

When the bolus has passed through the cricopharyngeus region, the esophageal phase of the swallow begins. The esophageal stage has greater variability in duration than the other stages of the swallow. Normal esophageal transit may vary from 8 to 20 seconds (Logemann, 1983). The upper third of the esophagus is composed of mixed voluntary and involuntary muscle, and the lower two thirds is composed entirely of involuntary muscle. The lower esophageal sphincter acts as a valve into the stomach. This valve, like the cricopharyngeus muscle or upper esophageal sphincter, must relax in time to allow the bolus to pass from the esophagus into the stomach.

In summary, the pharyngeal stage of the swallow is responsible for transit of material into the esophagus and for airway protection, probably the most important aspect of deglutition. Fig. 95-4 illustrates the progression of the bolus through the pharynx. If an anatomic or neuromuscular disorder affects the pharyngeal stage of the swallow, aspiration may result.

Aspiration

Aspiration is the entry of material into the airway below the true vocal cords. The effect of chronic aspiration in adults is not well understood but in the long term may result in slow pulmonary changes. Many head and neck cancer patients who suffer chronic aspiration do not survive long enough to experience long-term effects. Aspiration may occur before, during, or after the reflexive swallow (Logemann, 1983). Aspiration that occurs before the reflexive swallow may result from one of two disorders: reduced tongue control or a delayed or absent swallowing reflex. When tongue control is reduced during the oral preparatory or oral stages of the swallow, bits an pieces of food may fall into the pharynx and the airway as the patient chews. It is important to note that the airway is normally open during the oral preparatory and oral stages of the swallow. The airway is closed for only a fraction of a second during the reflexive or pharyngeal stage. Aspiration before the swallow can also occur because of a delayed or absent swallowing reflex. In this case the food is propelled out of the oral cavity by the tongue and falls into the pharynx, where it may come to rest in the valleculae, the piriform sinuses, or the airway *before* the reflex is triggered. The entry of food into the airway does not always trigger a cough reflex, particularly in neurologically impaired patients and some head and neck cancer patients in whom sensory input is damaged (Linden and Siebens, 1983; Logemann, 1983).

Aspiration *during* the swallow takes place when airway closure is inadequate to prevent material from entering the airway during the reflexive swallow. Aspiration *after* the swallow occurs for a number of reasons, including reduced laryngeal elevation, reduced pharyngeal peristalsis, unilateral pharyngeal paralysis, or cricopharyngeal dysfunction. In all of these instances, residue remains in the pharynx after the swallow. When the patient opens the larynx to inhale as is normal after a swallow, some of this residue is sucked or falls into the airway. Determination of the cause of aspiration in a particular patient is critical for efficient treatment. Aspiration is only a symptom of a disorder, and treatment is designed to eliminate the reason for the aspiration. Each of the disorders noted above as potential causes of aspiration has a different treatment or treatments.

Evaluation of Swallowing

Because aspiration occurs in many patients without visible external symptoms such as coughing, a careful evaluation of swallowing is imperative for patients who have been temporarily removed from oral feeding because of surgery or a neurologic incident such as a cerebrovascular accident of closed head trauma. Evaluation is particularly important at the time when oral feeding is being reinitiated. The currently available procedures to examine swallowing include manometry, ultrasonography, and fluoroscopy.

Manometry

Manometry is usually employed to examine esophageal peristalsis and the adequacy of functioning of the upper an lower esophageal sphincters (Fyke and Coe, 1955; Kelley, 1970; Van Trappen and Hellemans, 1967). The patients swallows a soft tube containing three pressure-sensitive gauges. The three sensors are positioned so that one registers the pressure in the cricopharyngeus muscle or upper esophageal segment, the second registers pressure changes in the body of the esophagus, and the third registers pressure in the lower esophageal sphincter. The gauges can also be positioned to register pressure in the pharynx, cricopharyngeus muscle, and body of the esophagus, thus permitting observation of the temporal relationship between pharyngeal peristalsis and opening or relaxation of the cricopharyngeus muscle. Manometry can give no information on aspiration or on the functioning in the oral cavity or larynx, which are critical components of the swallowing mechanism that require assessment (Hurwitz et al, 1975).

Ultrasonography

Ultrasonography is used to assess the anatomy and physiology of the tongue during swallowing but cannot be employed to examine the pharynx or larynx because of skeletal interference (Shawker et al, 1983). Thus it is limited to the study of the oral stage of deglutition.

Fluoroscopy

Fluoroscopy is the only available procedure that allows direct observation of the upper aerodigestive tract during all stages of deglutition, beginning with the oral preparatory phase and terminating with the esophageal phase. However, the standard barium swallow or "upper GI" procedure is inappropriate for head and neck surgical patients or neurologic patients who are having severe problems swallowing and are at risk for aspiration.

In the usual barium swallow procedure the patient is given a cap of barium and asked to swallow repeatedly. This procedure is designed to examine anatomy and motility in the esophagus, which is a collapsed tube that must be filled with barium to be viewed. For accurate evaluation of peristalsis without the effect of gravity the patient lies in a supine position and is usually viewed in the anteroposterior (AP) plane.

The modified barium swallow technique is used to examine the oral cavity and pharynx radiographically, which presents very different problems from those encountered when examining the esophagus. By definition the oral cavity and pharynx are cavities. When they are filled with large amounts of material, structure and function are obliterate rather than enhanced. In addition, the patient with an oral or pharyngeal swallowing disorder has greater difficulty handling a large bolus and is at much higher risk of aspiration. Thus modification of the barium swallow technique is necessary when assessing patients with oral and pharyngeal swallowing disorders. Only small amounts of material (generally approximately one-third teaspoon) are given per swallow. A variety of food consistencies is used, since patients may exhibit difficulty in swallowing liquids but not heavier food, or vice versa.

Another difference between the modified barium swallow technique and the traditional barium swallow is the patient's position. During the standard barium swallow the patient is lying down to counter the effects of gravity and facilitate evaluation of esophageal peristalsis. During the modified barium swallow the patient is seated upright in a normal eating position. In dysphagic patients the physiology of the oral cavity and pharynx in the normal eating position must be established first. Then, if desired, the patient's position can be changed to note postural effect on the drainage of food during swallowing. In the modified barium swallow procedure the patient is initially viewed in the lateral plane with the fluoroscopy tube focused on the lips anteriorly, the pharynx posteriorly, the soft palate superiorly, and the seventh cervical vertebra inferiorly as shown in Fig. 95-5. This allows complete assessment of oral and pharyngeal motility, as well as of aspiration in the pharyngeal aspect of the swallow. In the AP view the trachea overlies the esophagus, making assessment of aspiration difficult.

The major purposes of the modified barium swallow are (1) to define the oral and pharyngeal motility disorders occuring during the swallow; (2) to identify the presence of aspiration during swallows of any food consistency; (3) to determine which oral and pharyngeal motility disorders are causing aspiration so treatment for these disorders can be initiated; and (4) to assess the speed of the swallow (as compared with normal) to determine if the patient is receiving adequate nutrition by mouth.

Swallowing Disorders from Treatment for Head and Neck Cancer

Radiation therapy

Treatment for head and neck cancer often affects some aspect of deglutition. Radiation therapy for the oral cavity and pharynx, in addition to reducing salivary flow and causing usually short-term oral or pharyngeal edema, can lead to difficulties with triggering of the swallowing reflex and pharyngeal peristalsis that show up months after the therapy is finished

(Logemann, 1983). Presumably, the problem with pharyngeal peristalsis relates to increased fibrosis in the pharyngeal constrictors. The symptoms vary greatly in duration and severity from patient to patient and may appear 6 months or more after completion of a full course of radiation therapy. The long-term effects of radiation therapy on swallowing deserve greater attention and more thorough study.

Surgical treatment

The effects of surgical treatment for head and neck cancer on swallowing depend on the site and extent of the resection and the nature of reconstruction (Logemann, 1983; Logemann and Bytell, 1979). Some data indicate that the reconstruction will determine the swallowing pattern of the patient postoperatively, particularly when small amounts of tissue are resected.

Tongue

Within the oral cavity the tongue is the most important organ for maintaining normal food intake. The tongue is responsible for oral manipulation to prepare food for the swallow and is responsible for the oral initiation of the swallow. Thus every effort should be made to reconstruct the oral cavity to permit as much range and coordination of tongue motion as possible. Treatment for tumors in the anterior oral cavity may affect tongue mobility during chewing and the initiation of the swallow. In most cases of anterior oral cavity involvement, however, the reflexive swallow and the pharyngeal stage of deglutition are left unimpaired.

Posterior oral cavity

When a tumor invades the posterior oral cavity, a number of swallowing actions may be affected. These include tongue mobility during oral preparation and the oral initiation of the swallow, triggering of the swallowing reflex, velopharyngeal closure, and pharyngeal peristalsis. Surgery at the tonsil and base of the tongue directly invades the area where the swallowing reflex is triggered and also damages the attachments of the pharyngeal constrictors into the base of the tongue, thus reducing pharyngeal peristalsis. Reconstruction of this area with a thick and bulky flap interferes with functioning of the pharynx, tongue, and remaining normal tissues (McConnell and Teechgraeber, 1982). Utilization of the tongue in closure can significantly reduce the patients ability to handle a normal diet (Logemann, 1983).

Pharynx

Surgical involvement of the pharynx itself can significantly impair pharyngeal peristalsis. If the surgical resection involves only one side of the pharynx, the patient can usually be positioned to facilitate swallowing on the unoperated and less damages side. Turning the head to the damage side closes the piriform sinus on that side, thus directing food down the more normal side. Alternatively, the patient can tilt his head toward the more normal side to facilitate gravitational drainage of food down the stronger side. However, as in the oral cavity, the nature of the reconstruction may interfere with function on the unoperated side so that even with positioning the patient is permanently unable to handle food thicker than liquid and even has difficulty with liquid. Greater pharyngeal peristalsis is required to move heavier and thicker for food than is required for liquid.

Larynx

Resection of a part of the larynx, as occurs in hemilaryngectomy or supraglottic laryngectomy, can interfere with three aspects of the swallow: laryngeal elevation, laryngeal closure, and pharyngeal peristalsis.

In a hemilaryngectomy, laryngeal closure and occasionally pharyngeal peristalsis are the greatest problems. Depending again on the nature of the reconstruction and the nature and amount of tissue positioned on the operated side, many patients regain normal swallowing in a relatively short time (1 month) after hemilaryngectomy with swallowing retaining, which usually involves laryngeal adduction excercises (Jenkins et al, 1981).

Patients who have undergone supraglottic laryngectomy often have multiple problems with swallowing, which may include reductions in laryngeal elevation, laryngeal closure, and peristalsis. During the supraglottic laryngectomy procedure, portions of the hyoid bone are removed and thus laryngeal suspension from the hyoid is disturbed (Ogura et al, 1969; Shumrick and Keith, 1968). Unless the larynx is resuspended under the tongue, its lowered position in the neck creates a catchment area for food, which pharyngeal peristaltic action cannot clear away completely. The resultant residue at the top of the larynx may be aspirated after the swallow when the patient opens the larynx and inhales. Any damage to the true cords, the arytenoid cartilage, or their function in adduction will result in a sphincter that is incompetent to protect the airway an aspiration *during* the swallow. Removal of the supraglottic larynx also involves resection of a small part of the pharyngeal constrictors at their attachment to the larynx, and thus pharyngeal peristalsis may be affected so that residue is left in the pharynx (on the pharyngeal walls or in the piriform sinuses) after the swallow. The patient may inhale that residue and aspirate *after* the swallow.

As in the oral cavity, then, the extent of both resection and reconstruction is important in determining how quickly and well the partially laryngectomized patient can be rehabilitated in oral feeding. More research is needed to identify reconstruction techniques that result in the best deglutition postoperatively.

Posttreatment disorders

In addition to the disorders that result from various treatment procedures, posttreatment complications may arise that worsen the swallowing disorder. For example, after radiation therapy, osteoradionecrosis of the mandible may affect chewing. After surgical procedures, scar tissue may form or a fistula may develop that results in additional scar tissue, which further interferes with normal swallowing. Also, unexpected neurologic damage may result in paralysis of a vocal cord or part of the pharynx.

Summary

Aspiration is a generic term for a problem that has a number of specific causes. Treatment of aspiration in head and neck surgical patients requires accurate identification of the reason for the aspiration. Each of the known causes of aspiration (reduced tongue control, delayed swallowing reflex, reduced laryngeal closure, reduced laryngeal elevation, reduced pharyngeal peristalsis, and so on) requires a different treatment procedure. The best available procedures can be used to assess the patient's swallowing physiology, and the findings can be compared with normal data. Currently a fluoroscopic procedure, the modified barium swallow, gives the most valid results for the evaluation of swallowing physiology and the assessment of aspiration and its cause. The other techniques for assessment of swallowing, including bedside evaluation, manometry, and ultrasonography, cannot accurately assess aspiration. Studies have shown that the bedside evaluation of aspiration can be as much as 40% in error in identifying patients who are aspirating (Logemann, 1983). Optimum management of a patient with a swallowing disorder involves videofluoroscopic evaluation of swallowing at the time that oral feeding is to be reinitiated, so appropriate swallowing therapy can be planned, the patient can be quickly advanced from clear liquids to a normal diet, and the potential complications from chronic aspiration are avoided.

Many current treatment approaches for head and neck cancer compromise the function of the oral cavity, pharynx, and larynx. The extent of the dysfunction depends on (1) the extent of the surgical resection; (2) the nature of the surgical reconstruction; and (3) the appropriate and timely evaluation and treatment of the resulting swallowing disorder. The latter two factors, which are often more crucial, can be controlled to a great extent by the health care team. Whenever possible, the surgeon should attempt to reconstruct the affected structures using techniques that least restrict functional mobility.

A swallowing therapist can be involved in the patient's management from the time that head and neck cancer is diagnosed so timely evaluation and treatment of swallowing disorders can be initiated and as much normal function as possible can be restored in the shortest postoperative time.

Laryngeal Functions in Speech

Colin Painter

Studies of the way the larynx is used to produce phonetic contrasts in language may guide us in our search for an understanding of the details of larynx behavior when we generate speech.

Kenneth N. Stevens, 1981

In recent years a small but significant shift in emphasis has occurred in departments of otolaryngology in North America - a growing realization that laryngologists should be more than good surgeons and that part of their training should be an appreciation of the larynx as an organ of speech. It has been suggested that surgeons acquire a better understanding of the role of the speech pathologist an the singing coach (particularly, perhaps, of those who have expertise and experience equal to their own) and become more familiar with the work of the speech scientist. The assumption is that a physician who knows something about voice will better serve the voice patient.

In recognition of this trend, the editor of this volume has included sections on speech physiology and instrumentation for diagnosis and assessment. This discussion of the functions of the larynx in speech is intended as an editorial introduction to these sections.

Terra Incognita

Just as speech can be satisfactorily studied only in the context of language, disordered speech can be studied only by reference to normal. It is surprising how much is still unknown about the normal larynx.

In the realm of neural control in humans, we still lack details of (1) the somatic efferent nerve supply to the larynx and, in particular, nerve cell density and diameter in the pathways and terminal branching; (2) the maintenance of tonic postural activity by stretch receptors in the laryngeal muscles; (3) the phasic alterations in tone produced by reflex systems operated by receptors in the laryngeal joints; (4) the phasic alterations produced by reflex systems operated by pressure receptors in laryngeal tissue; (5) fibers of the autonomic system to the larynx; and (6) cerebellar control of laryngeal timing.

In the realm of laryngeal muscles, we still lack details for humans of the biochemistry of laryngeal muscle fiber types an the mass and stress-strain properties of the laryngeal muscles. Full-series whole-organ histologic studies of the larynx in all three planes are still not readily available, the aryepiglottic folds are known only in the most general sense, and the pharyngeal constrictors have not been adequately studied.

In addition, further study of the elastic properties of the laryngeal cartilages, the blood vessel network within the larynx, the surface microstructures of the laryngeal mucosa, and vocal fold vibratory patterns and how the air above the folds is disturbed in different ways for different voice qualities.

All of these are suitable topics for research. We have a tendency to take a mental set that says, "We do not, of course, know everything there is to know about the larynx, but it is just a matter of filling in the details". One wonders if this is indeed so. We might recall Fink's (1975) statement that "there is a natural propensity for the mind to recognize only the meaningful units of the visible and to ignore details it does not understand". There is no disagreement, however, about the basic function of the larynx.

Laryngeal Functions

The human larynx has evolve from four major an largely incompatible functions: airway opening for respiration and closure for protection of the respiratory system, effort closure for fixation of the trunk in strenuous activity, swallowing, and speech. In other words, how to stay alive an how to be *Homo sapiens.* Speech demands laryngeal configurations that are not too open an not too closed, somewhere between those adopted for respiration, particularly deep inspiration, and those found in effort closure, which are tighter even than the glottal stop. At the same time, speech is under voluntary control, whereas swallowing, with the rare exceptions of competition beer drinkers and sword swallowers, is a reflex.

The larynx has four primary functions in speech. (1) As a phonatory control mechanism it enables us to turn a direct flow of air from the lungs into an alternating on-off flow, which from an acoustic point of view is far more efficient. (2) As a fundamental frequency control mechanism it enables us to manipulate pitch in linguistically useful ways. (3) As an intensity control mechanism it enables us to use loudness levels contrastively. (4)

As a voice quality control mechanism it enables us to produce those subtle markers of geographic and social origin and those indices that mark the individual - age, sex, mood, and state of health.

Any examination of these four laryngeal functions in speech profits from a disposition to look both upstream and downstream. The linguist typically looks downstream from assumed causes (linguistic programming in the brain) to observable effects (laryngeal configurations) and may in the process neglect important laryngeal gestures. The physician typically works upstream by observing effects and then trying to explain them, and the dangers in working outside a theory are well documented. A strong case can be made for looking at both linguistic aspects and production mechanisms.

Larynx and Phonation

Linguistic aspects

Linguists have long recognized a primary distinction between sounds whose production is accompanied by vibrating vocal folds (voiced) and those that are not (voiceless) (Ladefoged, 1975). The four major types of sound in languages are vowels, vowel-like consonants, plosives, and fricatives. Vowels are almost always voiced. Certain consonants such as /m/, /n/, /l/, /r/. /w/, and /y/ have a consonantal function but are vowel-like from an acoustic standpoint and are also usually voiced. Plosives such as /p/-/b/, /t/-/d/, and /k/-/g/ are produced by bringing an upper and a lower articulator completely together (the upper and lower lips in the case of /p/ and /b/). This gesture obstructs the flow of air from the lungs so that pressure builds up behind the stop. The stop is later released, causing an audible explosion. Plosives are typically found as voiced (/b/, /d/, /g/) and voiceless (/p/, /t/, /k/) pairs at each place of articulation (point of maximum constriction in the vocal tract). Fricatives such as /f/-/v/ and /s/-/z/ are produced by bringing an upper and a lower articulator close enough to each other that air from the lungs is forced through a narrow gap. This results in turbulent flow at the place of articulation. Fricatives also are typically found as voiced (/v/, /z/) and voiceless (/f/, /s/) pairs. One linguistic feature, voicing, enables us to almost double the inventory of distinctive sounds in a language.

Linguists have also looked at the time course of voicing (Lisker and Abramson, 1964). For most speakers of English the utterance "a pea" is produced with nonvibrating vocal folds from the stop at the end of the first vowel until well after the release of the /p/, while "a bee" requires nonvibrating folds only from the stop to the release. On the other hand, for most speakers of French the /p/ of "un peau" (œ po) is produced rather like the English /b/, while the /b/, of "un beau" (œ bo) has voicing throughout. English /p/ equals French /p/ to a first approximation. Languages choose along a continuum of possibilities for their distinctions.

Other linguists (Halle and Stevens, 1971) have found themselves forced to talk not in terms of a binary pair on a one-dimensional continuum but by reference to a two- or three-dimensional matrix. For example, a language with a four-system of plosives at the lips, such as Hindi with /p/-/ph/-/b/-/bh/, must be described as both voiced/voiceless and aspirated/unaspirated, both dimensions being realized by laryngeal gestures.

Production mechanisms

The almost universally accepted myoelastic-aerodynamic theory of phonation was put into its modern form by Johannes Müller in 1848 an was presented again by Janwillem van den Berg in 1958. It traditionally describes one cycle of vibration of the vocal folds as follows. (1) The vocal folds are adducted to within 3 mm of each other by the action of the lateral cricoarytenoid and interarytenoid muscles. (2) Air is forced through the vocal tract from the lungs. (3) The folds are sucked together in accordance with Bernoulli's aerodynamic law. That is, given a constant volume velocity (air flow measured in cc/sec, U) from the lungs, the particle velocity must increase at a constriction such as the vocal folds, and in the process the air pressure (P) between the folds decreases, sucking the folds together. This can be demonstrated by hanging two balloons side by side and blowing between them; they are sucked together, not blown apart. (4) When the folds have been sucked together, the flow of air from the lung continues but the flow through the glottis (the space between the folds) ceases. The air pressure beneath the folds rises. (5) When the subglottal pressure is greater than the medial compression of the folds, the folds are blown apart (lateral displacement) and a puff of air is released into the supraglottal cavities. The glottal resistance to airflow may be calculated by dividing the subglottal pressure by the airflow. (6) The subglottal pressure falls. (7) Since the folds have been put in an almost adducted position by the adductor muscles, they seek to return to their position at the beginning of the cycle. The tissue elasticity of the folds accomplishes this. Elasticity may be calculated by dividing the restoring forces by the deforming forces. (8) A second cycle begins.

We are therefore in the presence of four forces (Lieberman, 1967), two muscle and two air: muscle tension and muscle elasticity, an subglottal pressure and intraglottal particle velocity. Each plays its part in one cycle of vibration of the vocal folds.

Certain preconditions are necessary for voicing. Van den Berg (1958) and Halle and Stevens (1971) have shown that the transglottal pressure (subglottal pressure divided by supraglottal pressure) and the airflow must be high enough, the glottal width small enough, and the glottal resistance not too great. The folds vibrate most easily when they are slack and approximated in the presence of a large pressure drop across the glottis.

More recent research (Stevens and Hirano, 1981) has revealed the crucial role in vibration of the vocal fold cover rather than the underlying muscle. Some authors have cast doubt on the importance of Bernoulli's law, and others have even challenged long-held views on the role of the intrinsic musculature in abduction and adduction. Fing (1975) has suggested that the elasticity of the laryngeal cartilages plays an important role.

Larynx and Fundamental Frequency

Linguistic aspects

The number of cycles of vibration of the vocal folds per second (called "fundamental frequency" or "f0" as a physical phenomenon, "pitch" when viewed perceptually, and "tone" as a linguistic unit) is used in a number of ways in language systems (Lehiste, 1970; Painter, 1979).

In a majority of the languages of the world, particularly those in sub-Saharan Africa and East and Southeast Asia and the Indian languages of the American continents, the pitch on which a syllable is uttered is as important as the consonants and vowel. A high-pitched "bá" might mean one thing and a low-pitched "bà" another. Usually there are only two contrasting pitch levels, high and low, although three levels are not uncommon. Asian tone languages also have glides from high to low or low to high or even two-directional glides.

The more familiar languages of Europe are not of this kind but do use contrastive series of pitches ("intonation contour"). This expresses the speaker's attitude to the world, as well as grammatical contrasts. Compare a grumpy "Good morning" pronounced "low/low/slightly higher" with a cheerful "low/high-to-low/low", or the statement "He's come" pronounced "low/high-to-low" with the question "low/low-to-high".

We also use pitch for stress to make a syllable stand out from its neighbors. Stressed syllables are also longer and louder, but pitch is the principal perceptual cue. Thus in the sentence "He pláced the phótogràph in the trây", pronounced "low/high/high/mid/mid/mid/mid/mid/low falling", one sentence stress (tray) moves, is different from the pitch in front of it, and has full vowel quality. Two additional primary word stresses (placed and phot-) have pitch change and full vowel quality but only level pitch. One secondary word stress (-graph) has full vowel quality but no pitch change and only level pitch. Five unstressed syllables are brief, indistinct in quality, not loud, and similar in pitch to the preceding syllable. Many pairs of words such as "súbject" and "subjéct" have such a primary word stress on the first syllable of the noun but the second syllable of the verb.

An individual may use a wide pitch range, as when excited or expressive, or a narrow one, as when depressed or in parkinsonism. A speaker's register (bailiwick) may be relatively high, as in puberphonia, or relatively low, like that of many newscasters. Singers, of course, have particularly fine control over the pitches used. Many motor speech disorders (nerve system lesions with dysarthria an dysphonia) (Darley et al, 1975) such as pseudobulbar palsy are described as having "monopitch". Absolute monopitch is probably never found, although the parson who substitutes length for pitch ("and the glooory of the Looord ...") to overcome the acoustics of a vaulted church is a possible example.

Production mechanisms

Pitch is controlled in one of two ways, by laryngeal muscle tension or by subglottic pressure, although the former is the more important. Laryngeal muscle forces are, in their turn, exerted in one of three directions: longitudinal, mediolateral, and vertical.

A long violin string has a lower fundamental frequency than a short one, but long vocal folds are associated with high pitch. This is because length is secondary to tension, and tense folds vibrate more rapidly, as do tense strings. Long folds also have less mass per unit length, and thin folds, like thin strings, have a high frequency of vibration. There are two ways to increase longitudinal tension, one internal and one external to the folds. The cricothyroid muscle is the principal pitch-raising muscle and lengthens the vocal folds by rotating the thyroid and cricoid cartilages around their joint. Conversely, the thyroarytenoid muscle shortens the folds but makes them more tense. We assume in these cases that the posterior cricoarytenoid muscle is acting to hold the arytenoid muscles back, and since the posterior cricoarytenoid is an abductor muscle, the interarytenoid muscles must ensure that the folds remain adducted. Clearly a delicate balance of forces is needed between these four muscles.

Medial compression is achieved by the adductor muscles, the lateral cricoarytenoid and the interarytenoid. Since the arytenoids are sitting on a sloping cricoid border, adduction by the lateral cricoarytenoid muscle tilts the folds down and in, which is difficult to visualize from above. Too much medial compression makes it difficult or impossible for the folds to vibrate at all. So does too little adduction, but within the vibratory range additional mediolateral tension increases fundamental frequency by raising glottic resistance and thus subglottic pressure during the closed phase of the glottic cycle.

Vertical tension is achieved by the extrinsic muscles of the larynx, that long chain running from the mandible through the hyoid to the sternum. The principal pitch-lowering muscle is the sternothyroid, and the largest of the extrinsic muscles, the genioglossus (Painter, 1976), seems to play an important role in raising pitch. The latter effect seems to be due to a stiffening of the vocal folds caused by the stretching of the conus elasticus as the larynx is raised. The untrained singer usually lowers the larynx for the low notes and raises it for the high ones.

An increase in subglottic pressure of 6 or 7 cm H_2O, either directly from the lungs or indirectly from increased glottic resistance, raises pitch by about half an octave. The physiologic constraints of the respiratory system therefore affect pitch. For this reason pitch usually falls at the end of a sentence, since not as much breath is left. We can, of course, ask the chest to do extra work when we want to raise the pitch at the end of a sentence for a question or put extra stress on a word.

At the very high end of our pitch range we may change the mode of vibration of the vocal folds by making them very long, stiff, and thin. In this falsetto voice, pitch is controlled by regulating the length of the vocal folds actually in vibration. Vibration is usually confined to the front part of the folds as pitch is raised in falsetto.

At the very low end of our pitch range we may change to yet another mode of vibration to produce creaky voice or vocal fry. The folds are tightly approximated but short, massive, and loose along their borders so the vocal fold cover vibrates slowly, only 20 to 80 times per second.

The normal adult male range spans the two octaves from 65 to 260 Hz with a mode at 130 Hz, one octave being a doubling of the fundamental frequency. Adult female voices are an octave higher, spanning the range from 130 to 520 Hz with a mode at 260 Hz. Singers' registers, however, are not only a matter of pitch; although the baritone cannot hit the tenor's high C except in falsetto, we categorize him almost as much for his overall quality.

Larynx and Intensity

Linguistic aspects

Pitch change as the principal cue for word and sentence stress is usually accompanied in English by increased loudness, although the latter may play a different role in other languages (Lehiste, 1970; Painter, 1979). In Czech, for example, all words are stressed on the first syllable. Since long and short vowels contrast in that language, duration is not available for prominence, and intensity plays an important role. In Polish, words are stressed on the penultimate syllable, in French on the last. In English and Russian the stressed syllables vary and the vowels in unstressed syllables change quality and are of less intensity. Compare the first syllables of "phótograph", "photógrapher", and "photográphic".

As with pitch, an individual may have a wide or narrow intensity range; narrow ranges are associated with dull dispositions and certain disorders such as amyotrophic lateral sclerosis. Overall intensity may also vary; loud voices are associated with aggressive personalities and soft ones with nonaggressive character. Persons with some disorders, such as chorea, exhibit unusual loudness variation.

Production mechanisms

The mechanisms for intensity change are essentially the same as those for pitch change, although intensity and pitch can be controlled individually. Intensity may be increased by greater laryngeal muscle tension, a longer closed phase, greater subglottic pressure, and greater airflow from the lungs.

An increase in glottic resistance caused by greater laryngeal muscle tension results in higher intensity for low and medium pitch utterances. At high pitch, however, the folds are already so tense that increased intensity has to be produced by increased push from the lungs. The primary result of increased vocal fold tension is a lengthening of the closed phase of the glottic cycle, although again, this is more true for low- and medium-pitched utterances. A longer closed period will result in greater short-term subglottic pressure, that is, a higher peak pressure within each glottic cycle, which in turn increases the intensity. Greater subglottic pressure and intensity can, of course, also be caused by an increased push from the lungs. Finally, we may say that greater airflow through the glottis results from either increased work done by the respiratory system or a larger glottic area. Ask an unsuspecting colleague to stand, close his eyes, and utter a prolonged "ah". Push him in the abdomen and the pitch and intensity will rise.

In general, greater intensity results from either increased work done by the respiratory system at constant vocal fold resistance or increased vocal fold resistance at constant airflow. When "ah", "ee", and "oo" are uttered with equal subglottic pressure, "ah" has more acoustic energy than the other two. However, with equal subglottic pressure all three are perceived as equal in loudness. Further, when they are uttered with equal acoustic energy (sound pressure level beyond the lips), "ah" is perceived as less loud. These three statements show that the perceived loudness of vowels largely depends on the rate of work done on the air in phonation and the physiologic effort needed to produce them.

The literature remains inconsistent on the correlation between intensity and lateral displacement of the vocal folds. Probably, other things being equal, greater vocal fold resistance reduces lateral displacement and hence glottic area in high-intensity utterances.

In one standard diagnostic technique the subject utters vowels at high, middle, and low pitch, each one at high, middle, and low intensity. The resulting nine-cell matrix demonstrates that intensity and frequency can be individually controlled even though they usually move in the same direction in speech.

Whisper typically has an intensity of 40 decibels (dB), normal female speech 65 dB, normal male speech 75 dB, and loud speech or shouting up to 90 dB.

Larynx and Voice Quality

Linguistic aspects

Whereas the voiced/voiceless opposition and changes in pitch and loudness are integral parts of all language systems, changes of voice quality are usually paralinguistic or extralinguistic features (Laver, 1980). That is, they are by the side of or out of the language system. What mood is the speaker in? How do we know it is Uncle Jim at the other end of the telephone? What is the speaker's sex, age, and state of health? From where does the speaker come? Do we recognize from the voice markers of socioeconomic status? The study of voice disorders (Aronson, 1980) fits within this semiotic framework with its multitude of subtle, an so far inadequately studied, indices.

Speech markers of the kind just described are the result of vocal tract size and shape, articulatory settings, secondary articulations, and configurations of the laryngeal vestibule and vocal folds. All speech events are dependent on vocal tract size and shape, but extrasystemic factors such as sex and age are obviously so. By "articulatory settings" I refer to the fact that a given language, dialect, or individual may be characterized by a certain center of gravity of the organs of speech from which particular sounds deviate. For example, in the Liverpool (Beatles') dialect, any given vowel is pronounced with the tongue more than normally up and back, whereas baby talk adopts a raised and fronted tongue. An upper-class British stiff upper lip may be accompanied by a stiff lower jaw.

Secondary articulations may be either systemic or extra-systemic. Whereas English has only two plosives made by raising the tongue to the alveolar ridge behind the teeth (/t/ and /d/), some other languages have four. Some simultaneously spread or round the lips, giving a four-term system /t/-/tw/-/d/-/dw/. Some simultaneously place the tongue in the position it occupies for the vowel /i/ = "ee", giving a four-term system /t/-/ty/-/d/-/dy/. Others simultaneously raise and retract the tongue, constrict the pharynx, or make glottic adjustments. These are called "secondary articulations", and all are found in addition extra-systemically in speakers of English.

However, as much as anything the configurations of the laryngeal vestibule and the vocal folds characterize a given voice quality.

Production mechanisms

An examination of photographs of the larynx taken during the utterance of vowels accompanied by different voice qualities shows that one can set up a continuum of "states of the glottis" or "phonation types". The glottal stop, tense voice, creaky voice, and normal voice all have the vocal folds and arytenoid cartilages approximated. A failure (or an intention) to bring the posterior edges of the arytenoid cartilages together leaves a glottic chink that adds a breathy component. Failure to bring the vocal processes together too leads to a glottic chink throughout the length of the glottis and adds a stronger breathy component. However, breathy voice is still voiced; although the vocal folds do not touch, they do flap in the breeze. When a small gap exists between the folds and a larger gap between the arytenoid cartilages, the folds do not vibrate at all, particularly when they are stiff, and the conditions are set up for whisper. A large gap between both the folds and the arytenoid cartilages also precludes vocal fold vibration and results in voicelessness or aspiration after the release of a plosive, but even this state of the glottis is not nearly as wide as in deep inspiration. Such a description of states of the glottis can handle most linguistic types and even many voice disorders, but since it is a one-dimensional display, it fails to capture some important facts.

A two-dimensional matrix does better (Halle and Stevens, 1971). It is possible, for example, to categorize qualities by both vocal fold width ranging from wide (as in the /ph/ of English "pot") to narrow (as in a choked "ahgh") and vocal fold tension ranging from slack (as in the /b/ of English "boy") to tense (as in the choked "ahgh" that is both tense and narrow). A three-by-three matrix of this kind enables description of nine voice qualities.

Even better is a three-dimensional matrix that in addition takes into account the cross-sectional area of the laryngeal vestibule. As xeroradiograms show clearly, a breathy voiced vowel is accompanied by a wide vestibule, whereas tense vowels are produced with a narrow vestibule. Of interest, although we still have rather little evidence, is the fact that vestibule area, glottic width, and glottic tension seem to be independently controllable. Speakers have a choice among at least three by three by three, or 27, voice qualities.

The same muscles we have described for pitch and intensity control glottic width and glottic tension. Vestibule area involves also the aryepiglottic muscle and possibly the pharyngeal constrictors.

Summary

What the physician typically wants from the speech scientist is a diagnostic protocol (Hirano, 1981), but such protocols will not be forthcoming until considerably more basic research, both acoustic and physiologic, has been done (Ludlow and Hart, 1981). Further, disordered speech and in particular voice problems can be studied only against the background of normal speech, and even there we have much to learn. It frequently happens that we look at a larynx that prompts us to write "Normal in appearance but sounds bad - send to the speech pathologist". This is as it should be, but when a normal-appearing larynx fails to make the appropriate gestures for neurogenic, psychogenic, or pulmonary reasons, or even from some organic cause that is not easily visualized, it would obviously be advantageous for the patient to be seeing a physician who has good intuition about how the larynx works. Laryngologists of this kind are the ones who have already acquired an understanding of the work of the speech pathologist, the singing coach, and the speech scientist.

Laryngeal Afferent and Efferent Systems in Voice Production

John A. Kirchner

The human larynx serves two main functions: protective and phonatory. The protective closure of the larynx excludes everything but air from the lungs and operates on a readily demonstrable reflex system.

The phonatory function, on the other hand, is monitored and controlled to a large degree by the auditory apparatus, as evidenced by the flat, unmodulated voice of the completely deaf. Whether the auditory apparatus alone is responsible for the high degree of coordination between the central nervous system, respiratory muscles, extralaryngeal, pharyngeal, and oral musculature and the laryngeal muscles themselves during singing and speech is, however, a question that has not been fully answered.

How, for an example, can a singer produce a precise pitch at the moment the sound is emitted? Is there a receptor system within the various components of the vocal tract that signals position and tension within the vocal cords and that is therefore independent of any additional adjusting influence triggered by the hearing mechanism, once the sound has been admitted? This latter concept gains credence from the fact the electromyographic activity in the cricothyroid and thyroarytenoid muscles begins before the sound is produced (Buchthal and Faaborg-Andersen, 1964; Hirano et al, 1970).

The initial "presetting" of the vocal apparatus, even if it occurs at the cortical level (Penfield and Roberts, 1959), depends at least in part on information originating in mechanoreceptors within the pharynx, tongue, facial structures, thorax, and abdomen, which help the performer place the laryngeal structures into a position that past experience leads him or her to believe will produce the desired sound once it is emitted.

To support the thesis of an afferent system originating within the larynx and participating in reflex control of vocal performance requires first that appropriate mechanoreceptors be identified in those parts of the larynx that are subjected to changes in pressure, tension, and length during phonation. Second, afferent discharges should be recordable from the parent nerves. Third, electrical stimulation of these receptors or their afferent nerve should evoke a response in the motor nerves to the vocal cord musculature.

Having established this type of reflex arc, the investigator is still left with the unanswered question of how these reflexes affect voice production. It is unlikely that a controlled study can be designed for human subjects or for experimental animals. Yet the evidence now available for an afferent contribution to phonation is impressive.

Mechanoreceptors

Subglottic trachea

Although the greatest population of receptors within the laryngeal mucosa serves a protective closure function and can be inactivated by topical anesthesia (Konig and von Leden, 1961a), a relatively dense population of mechanoreceptors exist in the subglottic trachea (Van Michel, 1963). Their threshold is such that they are certainly capable of stimulation by the subglottic pressures that occur during speaking and even more during singing (Bartlett et al, 1976; Sampson and Eyzaguirre, 1964).

Larynx

Muscle studies and spiral nerve endings coiled around individual muscle fibers have been identified in small numbers in the larynx of animals and humans (Abo-El-Enein and Wyke, 1966; Konig and von Leden, 1961a, 1961b; Lucas-Keene, 1961). Moreover, the presence of large afferent fibers (10 to 15 microm) in the recurrent laryngeal nerve, making up 2% of its fiber population, suggests the possibility of a stretch-sensitive mechanoreceptor system in the laryngeal muscles (Gacek and Lyon, 1963). Attempts to demonstrate that these structures can be excited during passive stretch of the muscle have usually yielded negative or equivocal results (Hirose, 1961). Wyke (1974), however, has reported increased motor unit activity in an intrinsic laryngeal muscle of the cat during passive stretch.

Electrophysiologic Recording of Discharges

The second source of evidence for an afferent contribution to reflex modulation of voice is provided by electrophysiologic recording of discharges arising in response to excitation of receptors by stimuli that occur during phonation. These include vibration, touch, pressure, passage of air across mucosal surfaces, and changes in the length of the vocal cords (Kirchner and Suzuki, 1968). These discharges have been recorded from the superior laryngeal nerve for the supraglottic larynx and from fibers within the recurrent laryngeal nerve from the subglottic larynx (Fig. 95-6) (Suzuki and Kirchner, 1969).

Feedback Mechanism

An additional source of afferent information has been identified in the fibrous capsules of the cricothyroid and cricoarytenoid joints, where corpuscular nerve endings have been described (Jankovskaya, 1959; Gracheva, 1963). The presence of these structures provides substantial evidence for a feedback mechanism originating within the larynx, since similar receptors located within joint capsules of the extremities are known to make a major contribution not only to the perception of posture and movement but also to the reflex regulation of muscle tone in the legs and arms (Boyd, 1954; Ekholm et al, 1960).

Although corpuscular nerve endings had been identified in the cricothyroid joint capsule, the recording of discharges from these receptors eluded investigators for many years, until Suzuki and Kirchner (1968) found these discharges in fibers within the external branch of the superior laryngeal nerve whose function had always been regarded as purely motor. Afferent fibers within the external branch were found to carry impulses in response to light pressure, traction, and rotation applied to the capsule of the cricothyroid joint. These responses were consistently abolished by local or topical anesthesia applied to the joint capsule (Fig. 95-7).

Electrical stimulation of the central stump of the external branch of the superior laryngeal nerve consistently produces reflex-evoked responses in both recurrent laryngeal nerves. That mechanoreceptor fibers are responsible, however, cannot be assumed, because other afferent fibers in the same nerve trunk supply sensation to a small, triangular area of mucous membrane in the anterior subglottic larynx (Suzuki and Kirchner, 1968).

Afferent fibers in the recurrent laryngeal nerve supply the subglottic mucosa and presumably signal pressure variations during phonation. Electrical stimulation of the central stump of the recurrent laryngeal nerve after it has been cut produces an evoked reflex response in both the ipsilateral and contralateral recurrent laryngeal nerves (Suzuki and Kirchner, 1969).

Although a reflex system originating within the larynx and unrelated to that governing protective sphincter closure is supportable on experimental evidence, the contribution of such a mechanoreceptor system to reflex modulation of laryngeal muscle activity during phonation remains speculative.

Experimentation

Designing a controlled experiment in the one mammal capable of singing would be difficult. Doing so would require testing the various features of vocal quality such a range, volume, stability, jitter, and shimmer after local or topical anesthesia applied precisely and exclusively to those parts of the larynx in which receptors or their afferent nerves are located. Gould and Okamura (1974) and Tanabe et al (1975) studied the effect of laryngeal topical anesthesia on voice production. No influence was observable during sustained vowel phonation at normal pitch and intensity, but at high-pitched, loud phonation an increase in glottal resistance and subglottic pressure was observed. These results must be interpreted with caution, however, since the extent of surface anesthesia cannot be precisely delineated and the possibility of motor nerve blockade cannot be ruled out. A similar study by Horii and Weinberg (1979) proposed several other possible sources of error in interpreting the refinements of vocal performance under topical anesthesia. These included the lack of cough reflex, which allows the accumulation of liquid on the vocal cords; excessive mucus secretion or physical change in the tissues themselves in response to the anesthetic agent; and obviously the spread of topical anesthesia to the adjacent trachea, pharynx, or base of tongue. Atropine, usually administered to reduce salivation and dry the mucosal surfaces for more effective anesthesia, might also influence vocal quality.

Transcutaneous injection of an anesthetic agent into the cricothyroid joint capsule without some of the material spreading to the adjacent recurrent laryngeal nerve or cricothyroid muscle is probably impossible. The contribution of articular mechanoreceptors to voice quality must probably remain speculative, given the technical difficulties of such an investigation in human subjects.

Shultz-Coulon (1978), who masked performers' ears with white noise as high as 100 dB during vocal tasks, studied the acoustic monitoring mechanism and its effect on vocal performance. Significant differences were observed in variations in fundamental frequency and in "initial overswing" among the three test groups of trained singers, untrained singers, and dysphonic patients.

Summary

Insofar as function can be deduced from structure, the component parts of a feedback mechanism are present within the larynx itself. Mechanoreceptor end organs have been identified in those parts of the larynx that are subject to pressure, stretch, and vibration during speech and singing.

Afferent nerve fibers supplying these receptors have been identified and afferent discharges recorded from them during light touch or passage of air across mucosal surfaces, during pressure applied to the laryngeal joint capsules, and during stretch applied to the vocal cords.

Electrical stimulation of these afferent fibers evokes responses in the recurrent laryngeal nerves. How much these reflexes contribute to the quality of voice during sustained phonation, however, is not entirely clear, particularly since other feedback mechanisms have been identified in the intercostal and abdominal muscles (Bishop, 1973). Additional reflex mechanisms exist in the pharynx, tongue, and facial muscles, all of which probably contribute to a system of feedback control over the phonatory function of the larynx and over the rapid, precise, and continuous adjustments that occur within the organs of speech.

Vocal Folds and Vocal Fold Vibratory Patterns

Joel C. Kahane

Recent research on the structure and mechanics of the vocal folds has shown them to be far more complex than previously thought. The structure of the vocal folds suits them well for transducing aerodynamic forces and maintaining highly specific physical conditions during sound production.

Subtle alterations in structure or viscoelastic properties of the vocal folds appear to introduce significant disequilibrium into the system, resulting in dysphonia. A more thorough understanding of the structure and biomechanical properties of the vocal folds should be helpful in developing insight into normal dysphonic voice production.

Histology

Hirano's detailed light and electron microscopy studies (1974, 1975) of the human vocal folds have resulted in the development of a dramatically different perspective on vocal fold structure and function. Hirano has shown that the vocal folds are multilayered, consisting of an epithelial layer, three connective tissue layers, and the vocalis portion of the thyroarytenoid muscle. Each layer has distinct structural and mechanical properties that are essential to the functional integrity of the vocal folds as a vibrator. Fig. 95-8 schematically illustrates the layered structures of the vocal folds. A detailed discussion follows.

Epithelial layer

The epithelial layer of the vocal folds forms the outermost layer of the mucosa, which also includes the subepithelial connective tissues of the lamina propria (Fig. 95-8). Three types of epithelium cover the vocal folds (Fig. 95-9). The areas superior and inferior to the vibrating portion of the vocal folds are covered with typical respiratory epithelium (pseudostratified ciliated columnar) with goblet cells. The vibrating edge of the vocal folds is covered with nonkeratinized stratified squamous epithelium that is separated from respiratory epithelium by small zones of transitional columnar epithelium. The epithelium is anchored to underlying connective tissue via a thin basement membrane. According to Hirano (1981b), the epithelium in the middle of the membranous portion of the vocal fold is eight cell layers thick (mean 0.05 mm) and extends 4.1 mm in the vertical plane.

Subepithelial connective tissues

Beneath the epithelium and superficial to the musculature of the vocal folds lie three layers of connective tissue, called the *lamina propria.* Hirano (1981b) has shown that the three layers are distinguishable from one another based on the distribution of the fibrous components (that is, elastic and collagenous fibers). The superficial, intermediate, and deep layers of the lamina propria are described below.

Superficial layer

The superficial layer of the lamina propria is locate subadjacent to the epithelium and corresponds to Reinke's space. It has a mean thickness of 0.3 mm (range 0.2 to 0.5 mm) and contains scant amounts of loosely interwoven elastic and collagenous fibers (Figs. 95-8 and 95-10). The collagenous fibers are 0.5 to 0.7 microm in diameter.

Intermediate layer

The intermediate layer of the lamina propria lies deep to the superficial layer and histologically is readily distinguishable from it. It is composed mainly of branching elastic fibers that measure 0.5 to 1.5 microm in diameter (Fig. 95-8). Sparse amounts of collagenous fibers are present, but the elastic fibers compose the vast majority of this layer (Fig. 95-10). The intermediate layer is not readily separable from the fibers of the deep layer of the lamina propria, which together form the vocal ligament.

Deep layer

The deep layer of the lamina propria is composed largely of collagenous fibers (Fig. 95-10) that are densely packed, slightly twisted, and formed into bundles that course parallel to the edge of the vocal muscle (Fig. 95-8). The vocal ligament, which is formed by the elastic and collagenous fibers of the intermediate and deep layers, has a mean thickness of 0.8 microm (range 0.5 to 1.1 microm). Owing to its composition, the vocal ligament is capable of yielding to longitudinal forces, yet is still resilient.

Anterior and posterior thickening

The lamina propria is thickened anteriorly as it approaches its attachment to the deep surface of the thyroid cartilage and is called the *anterior macula flava.* In the region of the vocal process, the thickening of the lamina propria is called the *posterior macula flava.* These differences reflect regional variations in the layers of the lamina propria. The superficial layer of the lamina propria is thickest in the midportion and thinnest at the anterior and posterior points of attachment. The intermediate layer is thickest at the anterior and posterior aspects of the folds and thinnest in the midportion. The deep layer is thickest at its posterior attachment to the vocal process.

Vocalis fibers of thyroarytenoid muscle

The medial fibers of the thyroarytenoid muscle are often referred to as the "vocalis muscle", and these form the muscular portion of the vocal folds. Sonesson (1960), in a comprehensive study of the anatomy of the vocal folds, showed that muscle fibers from the thyroarytenoid muscle did not insert directly into the vocal ligament. Rather, the vocal ligament and vocalis muscle fibers run parallel. Sonesson (1960), Hirano (1974), and others have shown that elastic fibers from the conus elasticus intermingle with the connective tissue surrounding and investing the muscle. The relationship between conus elasticus and the vocalis muscle is shown in Fig. 95-11.

Blood supply

Mihashi et al (1981) conducted a comprehensive study of the vascular supply of the vocal folds using Softex contact microangiography. They found regional differences in the distribution of blood vessels in the vocal fold. Blood vessels in the free (vibrating) edge entered the vocal fold from either the anterior or posterior end and ran perpendicular to the longitudinal axis. These vessels were clearly distinguishable from vessels in the lamina propria, which are thicker and run perpendicular to the vessels in the mucosal surface. Vessels in the lamina propria exhibit branching and form rich arterial and venous networks. Blood vessels to the vocalis muscle fibers enter from the deep surface of the muscle, pass longitudinally, and exhibit great arborization.

Interestingly, Mihashi et al (1981) reported a distinct vascular pattern in the midportion of the membranous vocal fold. This area exhibits the greatest movement during vibration. Blood vessels were found to be sparsely distributed, with numerous branches and many arteriovenous anastomoses. This arrangement may reduce resistance to vibratory activity while providing sufficient vascular supply to dissipate heat generated during vocal fold vibration.

Epithelial Specialization and Vocal Fold Lubrication

Electromicroscopic studies (Tillmann et al, 1977) of human laryngeal epithelium have shown it to contain surface specialization, called *microridges.* These are fine ridges or plicae, derived from plasmalemmal folds, in the superficial cell layer of the epithelium. The tips are club shaped or branching (0.6 microm in width, 1.5 microm in height) and are arranged in two basic patterns: either into parallel arrays or into widening circular arrangements across the surface. Frequently, microridges are found intermingled with less populous microvilli that also are present. Microridges are thought (Sperry and Wasserung, 1976) to play important roles in facilitating the spread and retention of mucus along the epithelium. This is important to help to keep the epithelial surfaces moist and pliant, which contributes greatly to the mobility of the vocal fold cover during vibration and helps to prevent vocal fold abrasion resulting from contact during voice production and effort (sphincteric) closure. Fukuda et al (1988) designed an ingenious experiment using dogs to study the movement of mucus along the vocal fold during vocal fold vibration. They found that the lubricating fluid formed a bilateral column along the lateral side of the vocal fold. The fluid underwent rotation and moved perpendicular to the long axis of the fluid column. The authors found that a poorly lubricated larynx and vocal folds required very high airflow and pressure for sound production. They found that in dogs, inadequately lubricated vocal folds required approximately 25% more subglottal pressure and 50% more airflow to produce sound than adequately lubricated folds. These findings clearly underscore the functional importance of the epithelial adaptations present in the human larynx that facilitate the distribution and retention of seromucinous secretions on the vocal folds.

Biomechanical Properties

In addition to histologic differences, the layers of the vocal folds have been found to differ in mechanical properties. Hirano (1974, 1975) obtained load-strain data for each of the layers of the vocal fold in the coronal plane.

On a mechanical basis (that is, stiffness) the five layers of the vocal folds can be categorized into the *cover* (epithelium and superficial layer of lamina propria), *transition* (intermediate and deep layers of the lamina propria), and *body* (vocalis muscle fibers). The cover is the least stiff, the body most stiff, and transition intermediate. The ratio of stiffness for cover/transition/boy is 1:8:10. Thus the cover is inherently more mobile than the other portions of the vocal folds. The outer layers are more fluidlike than elastic. This is beneficial from the viewpoint of sound production because the cover of the vocal folds also forms the walls of the glottal cavity and is thereby readily able to influence the properties of the transglottic airflow. This effect must, of course, be viewed in conjunction with the geometry and volume of the rima glottidis. This point will be discussed more fully in subsequent portions of this chapter.

Hirano has also shown that longitudinal stress of the vocal folds in the anteroposterior direction, brought about by vocalis or cricothyroid muscle contraction, increases stiffness in the vocal folds by stretching elastic and collagenous fibers as well as the vocal muscle. Longitudinal stress increases stiffness along the anteroposterior plane of the layers opposite to that found naturally in their transverse dimension. When the vocal folds are lengthened, the greatest stiffness occurs in the cover, followed by the transition and then the body. Titze and Talkin (1979) point out that it is important to distinguish the contribution of transverse and longitudinal differences. They write that

> in the transverse direction, we may have an impedance matching phenomenon. The amount of aerodynamic energy coupled into the tissues is probably maximized by the gradually changing transverse properties. On the other hand, the vibrating frequency and vertical stability (of the vocal folds) is determined more by the longitudinal properties.

As mentioned previously, the lamina propria thickens at the anterior an posterior macula flava as a result of increases in the intermediate layer. Hirano (1981b) has suggested that these thickenings may cushion the vocal folds against stresses generated in sound propagation and also provide reinforcement for them at these maximally stressed areas.

Vocal folds as a viscoelastic system

Although the vocal folds are structurally heterogeneous, from a mechanical point of view they may be thought of as a viscoelastic continuum that is incompressible and spatially varying (Titze, 1973; Titze and Strong, 1975). Displacement of the vocal folds by aeromechanical forces result in transient distortions in vocal fold shape that vary in character along the vocal fold and occur at different times within a vibratory cycle. Although greatly influenced by subglottic pressure, these redistributions of vocal fold volume are highly dependent on the physical properties of the vocal folds and on the mechanical stress created by intrinsic laryngeal muscle activity.

Longitudinal stress along the anteroposterior dimension of the vocal folds is applied principally by cricothyroid or vocalis muscle contraction (Hirano, 1974, 1975; Titze and Talkin, 1979). During voice production this results in a decrease in mass per unit area of length and decreases in compliance and in transverse shearing in the vocal fol. Muscular activity increases the elasticity of the folds along the long axis by stretching collagenous and elastic fibers. This increases resistance to transglottic airflow and enhances elastic recoil forces, resulting in a greater rate of vibration per unit time and higher pitch.

Compliance and elasticity are inversely relate. Thus at a lower fundamental frequency, the cover of the vocal fold is lax and more compliant than at a higher frequency of vibration. The softer, more yielding surfaces of the compliant vocal folds vibrate more slowly, absorb greater amounts of subglottic pressure, and require greater effort to sustain oscillation.

The effects of increased vocal fold compliance are readily appreciated when the vocal folds collide during the vibratory cycle. At this time the compression of the folds increases and they remain approximated for a longer portion of each vibratory cycle. Less time becomes available in fundamental frequency and a lowering of pitch. Edema, polyps, cysts, nodules, and polypoid degeneration are conditions that increase the compliance of vocal fold surfaces.

Typical mechanisms for varying compliance of the cover of the vocal fold involve contraction of cricothyroid muscle, which thins the cover and decreases compliance. Contraction of the vocalis muscle shortens the body, makes the cover lax, and increases compliance. Activity of both cricothyroid and vocalis muscles produces a variety of compliance conditions that influence the quality of vocal fold vibration.

Prephonatory Laryngeal Adjustments and Vocal Attack

Faaborg-Andersen (1957, 1965) using electromyography demonstrated that the length, tension, mass per unit area, and glottal configuration were preset by as much as 0.5 seconds before phonation. These, of course, are mediated by adductor muscles, primarily the lateral cricoarytenoid, which acts on the membranous portion of the glottis and the interarytenoid muscles, which act to narrow the cartilaginous portion of the glottis. These muscular events accompanied by aerodynamic forces are referred to as the *prephonatory adjustment phase.* Immediately following it, about 50 to 100 msec before the initiation of expiratory airflow, the vocal folds begin their movement toward the midline. This is called the *phonatory attack phase.* Vocal fold vibration is typically observed before the vocal folds meet at the midline, although these oscillations are less regular than those of the remainder of the voicing effort. The Bernoulli effect appears to play a critical role in the initiation of vocal fold vibration, much more so than in maintaining it.

The rate and extent of vocal fold adduction are primarily responsible for the character of the attack phase (Koike, 1967; Koike et al, 1967). Three basic types of vocal attack have been recognized (Moore, 1938) and are referred to as simultaneous vocal attack, hard glottal attack, and breathy attack. Voice production initiated by *simultaneous vocal attack* is the "normal" mode of initiating phonation. It occurs with a gradual, well-regulated build up of subglottal pressure at the same time that the vocal folds make contact at midline. The onset of vocal fold vibration is relatively stable. With *hard glottal attack* (coupe de glotte), the vocal folds are firmly pressed together at the initiation of voicing. This results in necessarily higher levels of subglottal pressure usage to initiate vocalization, which is more effortful. *Breathy (or aspirate) attack* is associated with the delivering of egressive airflow before the vocal folds are completely adducted. Thus a turbulent signal is perceived before the vocal signal is produced. Moore showed that all speakers in the course of speaking use the variety of vocal attacks just described. The extend and frequency of their usage will be dictated by style, phonetic context, and the nature of the utterances produced.

Vibratory Patterns

Typical vibrating cycle

Research from excised larynges (Fukuda et al, 1983; Saito et al, 1981) has shown that the vibratory movements of the vocal folds are quite complex. The vocal folds vibrate in a somewhat elliptic manner, owing to the substantial amount of vertical displacement during vibration. The portions of the vocal folds that do not collide during each cycle (that is, those covered by pseudostratified ciliated columnar epithelium beneath the vibrating surface and along the superior surface of the fold) move primarily in a horizontal path, whereas the trajectory of the tissue along the medial surface is principally circular. The body of the vocal folds exhibits almost a totally vertical (up-and-down) motion. The temporospatial characteristics of the vocal folds during the vibratory cycle have been well documented over the years using a variety of methodologies, the most successful of which has been ultra-high-speed photography (Farnsworth, 1940; Hirano, 1974; Rubin and Hirt, 1960; Timcke et al, 1958). These studies have determined that the typical vibratory cycle consists of opening, closing, and closed phases. The rate and duration of each phase are influenced by the pitch and loudness used and by physical factors that will be discussed subsequently.

The typical representation of the three phases during a single vibratory cycle is shown in Fig. 95-12. During the *opening phase* the vocal folds are displaced laterally by subglottic pressure and the rima glottidis is opened. In the *closing phase,* elastic recoil and aerodynamic forces draw the vocal folds toward midline, and during the *closed phase* the folds remain approximated for a brief portion of the cycle.

The aforementioned phases of the vibrating cycle do not occur in the absence of aeromechanical events involving the vertical dimension of the vocal folds. Fig. 95-13 is a schematic illustrating vibrating activity in the vertical plane during one cycle of vibration (Hirano, 1981a). Hirano reconstructed these figures from frame-by-frame analysis of ultra-high-speed films of the vocal folds viewed from above.

When the vocal folds are approximated (1), impedance to airflow is high and subglottal pressure increases (2), the lower portions of the vocal folds are pushed aside. The upper portion of the vocal fold is pushed inward because of several actors: incompressibility of the folds, pressure changes within the glottis, and the mechanical linkage between upper and lower portions of the folds. The difference in timing between the opening of the lower and upper portions of the vocal folds is called *vertical phase difference* and is significant because it creates a wavelike movement of the vocal folds. The subglottic pressure head continues to move upward (3), progressively separating the upper edges of the vocal fold and pushing them outward (4-5). This is called the *opening phase.* A puff of air is released as the vocal folds are separated. As the vocal folds in their upper portion approach their maximum displacement, the lower portions that were displaced first begin to return toward midline (6-8). The upper edge follows slightly behind. These events occur during the *closing phase* and are mediated by elastic recoil and a sudden pressure drop between the vocal folds caused by the air streaming through the rima glottidis. The vocal folds open and close first along their lower edge. The restorative forces typically cause the vocal folds to remain in contact or a short time during the cycle (9-10); this interval is called the *closed phase.* Subglottic pressure builds up again, and the next cycle begins. The duration of the closed phase is greatly influenced by the extent of the adductor muscle activity, compliance of the vocal folds, and intensity of voice (that is, subglottic pressure). The characteristics of the vibratory cycle become altered with changes in pitch, loudness, and vocal register. These issues are addressed in the final section of this chapter and in Chapter 96.

Vertical phase differences and mucosal wave

Previous discussions have shown that the vocal folds do not vibrate as rigid units but exhibit a wavelike movement that consists of vertical as well as horizontal components. The vertical phase differences between movement of the lower and upper portions of the vocal folds are also accompanied by a traveling wave (Farnsworth, 1940; Matsushita, 1975) along the *mucosa,* which forms the lateral walls of the rima glottidis. This wave contributes greatly to the shape and texture of the glottis and plays an active role in shaping the airflow modulating the distribution of spectral energy from the larynx. Movement of the mucosal wave depends on a soft and pliant superficial lamina propria.

The mucosal wave is propagated in the vertical plane by the force of subglottic airflow. The mucosal wave extends for 5 to 10 mm (Stevens, 1977) along the vocal fold and moves at a rate between 0.5 and 1 m/sec (Baer, 1975; Hirano et al, 1980). Hirano et al (1980) have shown that the mucosal wave continues over onto the superior border of the vocal fold for a variable distance. The greatest amplitude of the wave is found in the region closest to the vibrating edge. The wave becomes significantly dampened at more lateral points because of the high surface tension of the mucosal surface.

Hirano (1975, 1981b) has emphasized the importance of the mucosal wave in normal voice production. He has shown that lesions, edema, and degenerative changes in the cover of the vocal folds have deleterious effects on the voice because of alterations in the mechanical properties and viscoelastic nature of the vocal folds.

Vocal stability

The vocal folds normally exhibit some degree of cycle-to-cycle variability in rate (fundamental frequency) and acoustic amplitude (sound pressure). Cycle-to-cycle variability in fundamental frequency is called *jitter* or *perturbation.* In general, mean cycle-to-cycle period differences less than 100 msec or variation less than 1% of the mean frequency is characteristic of normal sustained voice production. Excessive short-term frequency variability, often observed in pathologic voices, has been associated with the perception of harsh, hoarse, or rough voice quality (Bowler, 1964; Hillenbrand, 1988; Lieberman, 1963; Wendahl, 1963). Possible physiologic explanation for vocal jitter include asymmetry in structure and biomechanical properties of the vocal folds, unpredictable behavior of mucus on the vocal folds, disturbances in airflow (Fukuda et al, 1988), noisy integration of laryngeal motor unit activity causing uneven tension in the vocal folds (Baer, 1979, 1980), fluctuations in blood volume resulting from heartbeat or vascular pulse influences on the capillary beds of the vocal folds (Orlikoff and Baken, 1989a, 1989b), and the effects of normal aging (Linville and Fisher, 1985; Ramig and Ringel, 1983; Wilcox and Horii, 1980).

Cycle-to-cycle variation in vocal amplitude is called *vocal shimmer.* It is thought to reflect the regularity of glottal airway dynamics (Koike et al, 1977). Normal voice is characterized by a mean cycle-to-cycle amplitude difference of 0.7 dB or less. Mean shimmer tends to be higher in vocal pathologic conditions than in normal voice (Heiberger and Horii, 1982; Horii, 1980). The mechanisms causing shimmer are thought to be similar to those of jitter, although considerably less research data are available about them. Excessive vocal shimmer has been related to the perception of a "noisy" voice, but the roughness or hardness associated with it may be qualitatively different from those described for jitter (Hillenbrand, 1988).

The properties of the human voice are owed in great part to the unique structure and mechanical properties of the vocal folds. No other sound-producing animal has vocal folds so elegant. Much still needs to be learned about the vocal folds in health and disease. These endeavors will surely keep voice scientists and laryngologists busy.

Research in Laryngeal Physiology With Excised Larynges

Donald S. Cooper

Experimentation with excised larynges is not a subdivision of the subject matter of laryngeal physiology but a procedure for the study of voice production. Since circulatory, contractile, and neural effects are absent, the excised larynx is perhaps best regarded as an organic model of a larynx. It may be compared to the physical models that traditionally have been used in such fields as ship design, where models of ship hulls are tested in tanks with observation of patterns of fluid flow. In fact, keeping in mind the dependence of both fields on considerations of fluid mechanics, experimentation with the excised larynx may be regarded as similar to the testing of ship models, or of airplane models in wind tunnels, but with the geometry turned inside out. As with all models, there are appropriate techniques for such experimentation and limitations to the areas within which the analogy applies. However, because it combines use of the natural vibrating structure with a degree of experimental control that cannot be achieve in the living organism, the excised larynx has been an important preparation for the study of voice physiology. As Liskovius wrote in 1846 at the end of a sober consideration on acoustic basis.

As a consequence of this comparison and this correspondence of the tones of the dead and isolated human larynx with the tones of the living and unharmed human throat, experimentation on the dead larynx may indeed be used as one of the means for the investigation of the nature of the human voice.

The present discussion constitutes an overview of research in voice physiology with the excised larynx preparation. Although I do not expect that the survey can be comprehensive, I hope to provide a perspective on the development, techniques, achievements, and limitations of such research.

First Period: Leonardo to Liskovius

The first recorded experiments on voice production in which an excised larynx preparation was used appear to have been those of Leonardo da Vinci (1452-1519), as recorded in his notebooks. Pancocelli-Calzia (1943) has described Leonardo's studies in speech production in some detail, an more recently attention has been focused on Leonardo's research in fluid mechanics, where he displayed much originality. Experimentation with the excised larynx preparation constituted a point of convergence for these two intellectual currents. The relevant passages from his notebooks follow:

A means to ascertain how the voice is produced at the exit of the windpipe. One removes the windpipe and lungs of a man; if the lungs, filled with air, are rapidly compressed, one can immediately recognize how the pipe named the trachea produces the voice. One observes and hears this well also with the neck of a swan or goose, which people often make cry after it is dead (Anat. A, 3 r).

We will also arrange an experiment on it during the dissection of animals, by introducing air into their lungs and squeezing it out, while we narrow or widen the 'fistula', the producer of their voice (Wu An, IV, 10v).

It must be noted that Leonardo does not use the term *larynx* or any correspondingly specialized term, although he left drawings that clearly illustrate his awareness of its anatomy, but refers to it as *trachea,* and designates the glottis as *fistula.* Further passages are of substantial interest, particularly a brief discussion of the relationship between airflow in voice production and the geometric configuration of the larynx. No subsequent experimenter of record has used the lungs of the preparation as the air source. However, the basic components of such an experiment are given in Leonardo's notes: an air source, the trachea and larynx, control of the glottal configuration, and the dependent variable of interest, in his case the voice.

Leonardo's work did not become known until the twentieth century. Many different investigators have performed subsequent studies of phonation using the excised larynx preparation. Johannes Müller, in a discussion of voice physiology in his *Handbook of Human Physiology* (1837, vol 2, part 1), wrote, "Ferrein, Liscovius, and Lehfeldt have so far acquired the greatest merit for themselves in the theory of voice". The criteria for his selection are clear; this is a list of those who had published research on phonation based on excised larynges before Müller's own work. The French physician and anatomist Antoine Ferrein (1693-1769) presented his studies on phonation based on excised larynges to the Royal Academy of Sciences of France in 1741. The flavor of his work can be obtained from a brief extract in which he describes his first successful experiment. At first he was unsuccessful in exciting a human larynx by merely blowing into the trachea without controlling the glottal configuration. He wrote:

> I took a cadaver. I blew several times upwards into the trachea, the larynx was mute on this occasion.
>
> Subsequently I reflected that the voice requires not only a stronger wind, but also a new degree of narrowing in the larynx: I took that of a dog, I approximated the lips of the glottis, and I blew hard into the trachea: this time the organ seemed to come to life, and one heard, I do not say a sound, but a brilliant voice, more agreeable to me than the most touching concerts.

Ferrein carried out extensive experimentation on excised human larynges as well as on those of dogs, bulls, pigs, and sheep. He found that vocal intensity was controlled by glottal width and air velocity. Variation of airflow, he discovered, did not affect the fundamental frequency without simultaneous effects on intensity, while in the excised larynx fundamental frequency was controlled primarily by cord tension. Two other factors Ferrein considered as possible candidates for the control of fundamental frequency were the possible subdivision of the vibrating cords, which he considered impracticable, and the contraction of the vocal cords. He described his procedures in some detail (Cooper, 1989; Ferrein, 1746).

The subsequent work of Liskovius (1814, 1846), Lehfeldt (1835), and Müller (1837, 1839) brought such studies into the mainstream of physiologic study and established the broad outlines of the field of phonatory physiology. Basic questions, such as the mechanisms of frequency and intensity variation and the effect of the supraglottic portions of the larynx and upper vocal tract on fundamental frequency, were posed and answered, on the whole well. However, observable phonatory output quantities were limited to Fo and impressionistic judgments of voice intensity and quality.

The most famous experiments on the phonation of excised larynges, apart from van den Berg's, have been those of the great German physiologist Johannes Müller (1801-1858). Müller, professor of physiology at the University of Berlin and the teacher of several of the greatest physiologists of the nineteenth century, including Brücke, Helmholtz, du Bois-Reymond, and Ludwig, published the first section of his research in the second volume of his *Handbook of Human Physiology* (1833-1840), an encyclopedic work of more than three quarters of a million words (Boring, 1957). The second section of his research is contained in a small separate booklet, *On the Compensation of Physical Forces in the Human Organ of Voice* (1839). Although Müller's basic findings are embodied in the *Handbook*, his study of compensation is devoted to a quantitative study of the relationship between vocal cord tension and subglottal pressure in the control of fundamental frequency, with discussion of voice production in other mammals, birds, and amphibians. In this work Müller also used a different preparation, consisting of a cadaver head with removal of the cervical vertebrae. Adapting his techniques to evaluate the contribution of the upper resonant cavities to sound production, Müller was able to manipulate the lips and oral cavity to produce the vowels /u/ and /a/ and the consonants /m/ and /w/. Of the results he wrote:

> Hereby once more I did not become acquainted with any new elements, however the timbre of the voice becomes still more like that of a human being, and with the apparatus to be indicated, sometimes so similar, that any difference between the living body and the machine (the preparation) disappeared.

Fig. 95-14 illustrates the experimental apparatus used by J. Müller, and Fig. 95-15 shows a modern apparatus of van den Berg and Tan, which corresponds functionally to Müller's. As a source of appropriately warmed and humidified air, Müller and his contemporaries used their lungs, where a modern experimenter might use a stable blower and an air-conditioning device adapted from the area of respiratory therapy. Both the acoustic and aerodynamic characteristics of an artificial subglottal air system require systematic consideration (Baer, 1975; Ishizaka et al, 1976). The U-tube manometer gave Müller the ability to control average subglottal pressure. As Leonardo had already found, control of the glottal configuration was necessary, and in fact Müller and subsequent experimenters have found it possible to simulate the contraction of all intrinsic laryngeal muscles, either isotonically, for instance by weights on a small tray attached by a cord to the origin or insertion of the muscle in question, or isometrically, by a fixed attachment of the cord controlling the muscle. However, success in simulating the activity of the vocalis muscle has been limited, and later I shall refer to the results of this omission. Müller used a "compressorium", a metal construction to exert a bilateral medial force on the vocal folds, as have various subsequent experimenters, and he attached a cord to the thyroid cartilage to tip with respect to the cricoid in order to tense the vocal fold and raise pitch, as Ferrein had correctly noted, thus simulating the activity of the cricothyroid muscle.

Second Period: Harless to Nagel

Later studies from Müller up to World War I brought many advances. Realizing the limitation of the excised larynx experiment in terms of the absence of any adequate simulation of vocalis contraction, Harless (1853), Ewald (1898), and Nagel (1909) constructed models of the vocal folds using live muscle tissue that could be stimulated during phonation of a laryngeal model, a procedure that came to fruition in the work of E. Müller (1938) to be discussed later. Merkel (1863, 1866) and Grützner (1879) devoted studies on excised larynges to the detailed analysis of problems of register.

Although laryngeal aerodynamics has only recently become a well-developed field, during the nineteenth century problems in this area were being considered, primarily on the basis of laryngeal models or experiments with excised larynges. Although subglottal pressure provided an obvious source of opening force for the larynx, experimentation suggested that negative pressures arising from the Bernoulli principle provided a force to close the glottis in phonation. Harless (1853), a founder of laryngeal biophysics, expressed regret that the manometers of this time did not have the high-frequency response necessary to measure the dynamic variation of laryngeal air pressures during the glottic cycle, but he correctly hypothesized the presence of negative air pressures in the lower part of the glottis ("immediately below the cord") when the vocal folds were approaching each other at narrow glottal widths. Tonndorf's (1925) first explicit studies of the Bernoulli principle in phonation were anticipated on an experimental level by the finding of the Belgian Jesuit Ch. Looten (1870s) that in a phonating excised bovine larynx, in addition to regions of strong positive pressure, there were regions of strong negative pressure.

Other studies with excised larynges assisted in demarcating the area of the vibrating structures in voice production. As part of his theory of chest voice, Johannes Müller (1837) had written that "the cords vibrate with their whole width, so do the membranes connected to them and the thyroarytenoid muscle". French physiologists experimenting with the excised larynx preparation were responsible for demonstrating the importance of the mucosa in voice production. In 1866 Fournié described his finding that an excised larynx would no longer phonate if the mucosa was removed. Martel (1885) and Lermoyez (1886) concluded that the muscle of the vocal fold was not included in the vibrating portion at all; Lermoyez emphasized that, although both the vocal ligament and the superficial portion of the mucosa vibrated in chest voice (damage or removal of the ligament led to loss of voice, as Baer was to find again in 1975), in falsetto vibration was restricted to the superficial portion of the mucosa. These findings are close to the results of Saito and his group, which are described below (Fukuda et al, 1983; Saito et al, 1981).

Other studies for the first time made it possible to observe vocal fold vibration directly, first on a composite basis by the use of stroboscopy and then on a cycle-by-cycle basis. After Harless (1853) introduced stroboscopy to the study of vocal fold vibration, Koschlakoff (1886), and Réthi (1896) used stroboscopy to study the motion of the folds of excised larynges. Réthi used a microscope to observe the motion of bronze particles on the vocal fold of an excised larynx under stroboscopic illumination, essentially the same technique used by Baer. He provided a description of the vertical component of vocal fold movement.

In chest voice one sees the bronze particles located in the edges of the vocal folds in a large upward and downward movement, the farther forward and the farther upward, the smaller is the excursion, and at the height of the thyroid cartilage it is equal to zero.

Réthi also realized that the phenomena that earlier stroboscopic studies of Oertel (1878) and Koschlakoff (1886) had described as a "nodal line" on the vibrating fold probably constituted a stationary stroboscopic illusion corresponding to a single position of a mucosal wave propagated upward and laterally on the vocal fold during its vibratory cycle, which he described clearly.

Third Period: Weiss to Trendelendburg

An excised larynx study by Otto Weiss in 1914 initiated a new period of phonatory physiology by the publication of simultaneous cycle-to-cycle records not only of dynamic subglottal pressures, fulfilling a wish of Harless, but also of glottal width and sound (Fig. 95-16). Weiss (1914) recorded medial-lateral movement by shining a narrow beam of light through the glottis and a slit that crossed the folds at right angles (Fig. 95-17), resulting in an image recorded on moving film, which permitted separate observation of the movement of right and left vocal folds.

In the period between World Wars I and II, Wilhelm Trendelenburg and his students published a series of papers that made possible the resolution of vocal fold movement into simultaneous vertical and medial-lateral components. Like Johannes Müller, Trendelenburg was a physiologist of wide-ranging interests (Schütz, 1950). Probably it was partly the stimulation of his younger brother Ferdinand, a prominent acoustician, and partly his own interest in the physiology of musical performance that led him to produce a series of works on the physiology of speech production. Improved electroacoustic techniques suggested detailed analysis of the relations between the recorded acoustic signal and the vibratory pattern of the larynx (Trendelenburg and Hartmann, 1939; Trendelenburg, 1940). The pervasive contribution of Trendelenburg is indicated by the fact that almost all of the excised larynx studies listed for this period are his work or that of members of his group. Lullies (1953) summarizes much of this research; Trendelenburg (1942) provides a detailed overview.

A paper by Trendelenburg and Wullstein (1935) introduce a second measure of movement. A capacitive displacement transducer, with one capacitor plate suspended above or below the fold and the other constituted by the fold itself, was used to measure vocal fold movement simultaneously with the measurement of glottal width (Fig. 95-18). In the upper part of Fig. 95-18 we see records from a phonating excised larynx. The upper record is from the capacitive transducer and reflects vertical movement, but with some contamination from medial-lateral fold movement and the variation of fold thickness. The lower signal records the width of the glottis as a function of time. The complexity of the signal from the capacitive transducer led Hartmann (1938), another collaborator of Trendelenburg, to use optical techniques to measure the vertical movement of the vibrating vocal fold, suggesting a resolution of vocal fold movement into X and Y coordinates (Fig. 95-19); the earlier photographic techniques were refined by the use of a photoelectric cell by Wullstein (1936). However, with Hartmann's techniques the X coordinate refers to the location of the most medial tissue point in a given coronal section through the vocal fold, the identity of which varies over time. Similarly, the Y coordinate value refers to the location of the highest tissue point of the vocal fold in a coronal section, the identity of which also varies over time. The locations of these points would rarely coincide. Consequently the attempt by Trendelenburg's assistant E. Müller (1938) to combine these sets of values to plot a trajectory of vocal fold movement resulted in some distortion, a problem solved by the procedures of Réthi, Baer, and Saito in which the movement of a single point in a coronal section trajectory is tracked. Fig. 95-20 shows Saito et al's data (1985), in which the movement of a lead pellet inserted into the superficial layer of the vocal fold in a live dog is resolved into roughly sinusoidal X and Y components, the phase relations of which vary depending on the position of the particle. Kaneko et al (1981) have published similar results based on ultrasonography. Thus the general picture of vocal fold movement that emerged from the studies of Trendelenburg's school was similar to today's, but this group did not completely solve the technical problems of tracking vocal fold movement.

Fourth Period: Van Den Berg and Others

Apart from isolated observations, the excised larynx experiment was little used in the postwar period until van den Berg and Tan (1959a, 1959b; Tan, 1960) revived it. These studies are partially described in Tan's dissertation (1960). In a series of papers, van den Berg, a medical physicist, discussed his conclusions from these experiments, which brought systematic consideration of fluid mechanics into the study of phonation. The studies of van den Berg and his associates marked a new beginning in studies of the biophysics of voice production. Van den Berg applied modern instrumentation and analytic methods to the study of laryngeal vibration. In so doing, he established the myoelastic-aerodynamic theory of voice production, which emphasizes the interaction between the mechanics of the vibrating structures and the complex laryngeal airflow, in contrast to the neurochronaxic theory of Husson, who claimed that vocal fold vibration resulted from neural firing at the rate of the voice fundamental frequency. Since van den Berg's work touched almost every aspect of voice production, a detailed summary is beyond the scope of this discussion. The reader is referred to van den Berg's valuable survey of modern research in experimental phoniatrics (1962) for summaries and bibliography of his work. Von Leden (1961) and his colleagues used the excised larynx preparation in their high-speed photographic studies, refining the notions of fold movement that had emerged from studies of Trendelenburg's group. Later studies by Hiroto (1968) developed the modern concept of wavelike movement of the mucosal membrane in voice production, known as the mucoviscoelastic-aerodynamic theory of phonation. Anthony (1968), like Müller (1839), has used the excised larynx preparation to study voice production as part of a general physiologic model of speech production.

Modern Period

The landmark of the modern period of excised larynx studies has been Baer's dissertation (1975). Like Réthi (1896), Baer used a microscope to examine the movement of particles on the vibrating vocal folds of excised larynges under stroboscopic illumination. He found that particles on the surface of the vocal fold vibrated in roughly elliptic trajectories, with the orientation of the major axis of the ellipse dependent on the location of the particle (Fig. 95-21). The systematic treatment and integration of the findings with modern theories of the fluid mechanics of the larynx (Ishizaka and Matsudaira, 1972) make Baer's work a fundamental treatise on voice production, from which only part of the results has been published (Baer, 1973a, 1973b, 1974, 1981a, 1981b). Baer's study was completed almost simultaneously with the publication of Matsushita's (1969, 1975) optical study of phonation in the excised hemilarynx preparation, which permits direct observation of the medial aspect of the phonating larynx. The extent to which airflow patterns in the excised hemilarynx correspond to those in the full larynx remains to be determined.

The excised full larynx recently has come to be used as a testing ground for establishing techniques for use on human subjects. Holmer and Kitzing (1973, 1978) and Zagzebski et al (1983) carried out experimental studies of vocal fold movement by ultrasonography. Saito applied x-ray stroboscopy to the tracking of trajectories of lead particles injected into the vocal fold (Fukuda et al, 1983; Saito et al, 1981). Lecluse evaluated electroglottography in excised larynges (1977; Lecluse et al, 1974, 1975). The Iowa group used pressure transducers to study vocal fold contact stress and vocal process contact pressures (Titze and Scherer, 1981; Scherer et al, 1982, 1984, 1985).

A new and important direction has been the use of the excised larynx preparation to simulate vocal fold abnormalities, a procedure van den Berg suggested in 1962 and Shepherd (1974) carried through in his dissertation with regard to aerodynamic variables. In a similar vein, Isshiki (1977, 1981) has carried out studies of hoarseness and asymmetric fold tension on excised larynx preparations. Johannes Müller (1837) and Liskovius (1846) have also studied the effect of asymmetric fold tension in excised larynges. A related direction has been the use of the excised larynx preparation to specify the effects of specific forces or geometric changes produced by laryngeal muscles on the acoustic output (Kitajima et al, 1979). Such studies not only contribute to the understanding of the mechanics of the larynx, but also make the results of laryngeal surgery more predictable.

Some prospects and problems for the study of voice production by means o the excised larynx preparation deserve consideration. The influence of subglottal and supraglottal structures on phonation has become the subject of analytic treatment only relatively recently, but in various forms has been a repeated theme of studies with excised larynges. Ferrein wrote on the basis of his experiments that "one sees ... that the different tones are only the low and high sound of the vocal cords, for it is certain that the mouth and nose have no share in this change". Both Johannes Müller (1837) and E. Müller (1939) experimented with the effect of tubes of different lengths above and below the larynx, and Liskovius (1846) pointed out that in the absence of realistic subglottal and supraglottal resonances the excised larynx could not produce the same timbre as the human voice.

Attaching a supraglottal resonance tube to the excised larynx is difficult because of its irregular shape and the necessity to vary glottal configuration and have access for transducers. Trendelenburg and Wullstein (1935) carried out the most successful approach to this problem. They constructed resonant tubes modeling the configuration of the oral cavity in the production of vowels /a/ and /o/ but found no difference in vocal fold motion of an excised larynx using a capacitive movement transducer. In view of the renewed interest in this area and analytic advances with the work of Flanagan and Landgraf (168), Ishizaka and Matsudaira (1972), Rothenberg (1981, 1983), Titze (1983), Childers' et al (1983), and others, it may be useful to reinitiate experimental studies of the coupling between subglottal and supraglottal resonators and vocal fold movement by means of the excised larynx preparation, which with modern instrumentation may have much to contribute to solving the problem.

Finally, a limitation of the excised larynx should be mentioned. Two experiments cast light on the significance of the lack of adequate simulation of vocalis muscle contraction in the excised larynx experiment. To provide a contractile simulation of the vocal folds in a phonating model of the larynx, E. Müller (1938), a member of Trendelenburg's group, used two gastrocnemius muscles of frog, covered with a strip of fine rubber that simulated the mucosa. When a change in the internal tension of the fold was produced by stimulation of the muscles, little change occurred in the extent of vertical movement but the range of lateral movement decreased. Unfortunately, the interpretation of these results is made more difficult by the fact that the frog gastrocnemius muscle has a complex structure and differs notably in its fiber arrangement from the vocalis muscle. Fukuda et al (1983) compare the trajectories of lead particles injected into the vocal folds of live dogs with and without stimulation of the recurrent nerve. They found larger trajectories for the pellets in the mucosa when the muscle was activated. Both these findings suggest that the trajectories that would be observed in the vocal folds with activation of vocal fold muscles may differ in detail from those found with the muscle inactive, as in the excised larynx experiment. They also may eventually suggest procedures to compensate for this problem. This conclusion cautions that we must not assume that the excised larynx is a good model for phonation in all respects but must give consideration to what problems may best be approached in the excised larynx, what problems demand use of animal models, inorganic ones, or mixed models such as those of E. Müller, and what observations can be made only in human beings. One may also anticipate that in the near future intact viable larynges will be used for the investigation of voice production.

Role of Pharynx in Speech

Sidney Wood

Scholars of language, speech scientists, and speech therapists have always faced the problem of determining how speech maneuvers are related to the sounds heard? How does the vocal system work? What remedial advice can be given? The question is equally urgent for the otolaryngologist - head and neck surgeon: how will surgery affect the patient's speech?

Continuous-Transmission System

This discussion deals specifically with the role of the pharynx in speech production, but the underlying theory is relevant for the entire vocal tract (the supralaryngeal airways considered as a nonhomogeneous pipe in which sound is generated and filtered).

Zemlin (1968) has written of the pharynx:

The contributions of the pharynx to speech production are not fully understood ... Recently, however, thanks to the contributions of cinefluorography, it has been shown that some not particularly well-understood changes in the configurations of the pharynx do take place during speech production.

Remnants of an obsolete theory contribute to that lack of understanding. It is a deceptive oversimplification to divide the vocal tract into independent anterior and posterior cavities, each with its own natural resonance (the so-called mouth and throat resonances). The vocal tract should instead be seen as a continuous transmission system in which all parts contribute to every resonance (of which at least five are usually audible).

Nor is there any relevance in the position of the highest part of the tongue, a traditional reference point in phonetics; acoustically relevant maneuvers may be executed simultaneously in the buccal, velar, and pharyngeal regions. Models based on the highest point of the tongue neglect the pharynx. The earliest radiographic studies of speech from the turn of the century (Meyer, 1907; Scheier, 1909) confirmed that the pharynx narrows for "aw" and "ah" sounds (Fig. 95-22). Carmody (1941) reported detailed measurements of the pharyngeal cavity in speech, but he did not relate these details to a theoretic model. More recently the pharyngeal cavity has been explored in speech by tomography (Fant, 1960, 1965), ultrasonography (Minifie et al, 1970), and fibroscopy (Lindqvist and Sundberg, 1971). The extrinsic muscles of the tongue all contract in the pharyngeal region (Figs. 95-22 and 95-27).

The role of the extrinsic muscles in directing lingual maneuvers is well known (Fucci and Petrosino, 1981; Smith, 1971; Zemlin, 1968). The hyoglossi draw the tongue down and to the rear to narrow the lower pharynx; the posterior parts of the genioglossi pull the tongue root forward to widen the pharynx (and assist in raising the tongue body toward the palate); the glossopharyngei (fibers of the superior pharyngeal constrictors that insert into the sides of the tongue) draw the tongue back into the midpharynx; and the styloglossi draw the tongue up and to the rear toward the soft palate. The effect of contracting these muscles, alone or in combination, selectively aided by the intrinsic muscles of the tongue, the pharyngeal constrictors, the muscles of the palate and lips, and the mandibular depressors and elevators, is to widen or narrow the different regions of the vocal tract locally in order to tune its resonances. A type of model based on this behavior is compatible with observed motor activity and acoustic phenomena (Wood, 1979). In contrast, the relationships between the position of the highest part of the tongue, muscular activity, and vocal tract shape are ambiguous. Models based on this position provide a bewildering framework for interpreting speech maneuvers and the spectral output of speech.

Today we are in a better position to explain pharyngeal configurations by interpreting them with reference to acoustics theory, physiology, and their linguistic significance and, by implication, to foresee the impact of lesions on the patient's speech. The *source-filter theory* accounts for sound production in three stages: the generation of a complex sound at some source in the vocal tract, selective transmission of frequency components of this sound through the vocal tract, and radiation of the modified sound from the lips (Chiba and Kajiyama, 1941; Fant, 1967; Stevens and House, 1961). Resonance frequencies are usually found by searching for standing waves, formerly by sweeping an electric analogue of the vocal tract (Dunn, 1950; Fant, 1960; Stevens et al, 1953) currently by computer. In either case the configuration is represented by the vocal tract area function based on measurements obtained from lateral radiographic profiles. *Perturbation theory* relates a local change in vocal tract configuration to changes in the acoustic output by referring the modification to the sensitivity of resonance modes to local changes of vocal tract shape. The last is proportional to the local difference between kinetic and potential energy for each mode (Fant, 1980).

Most speech activity is concealed from view. Vocal tract configurations and internal speech maneuvers are therefore conveniently studied on midsagittal radiographs or cinefluorographic films (see the section of Chapter 96 by Painter). The first motion films were made with continuous irradiation, low camera speeds that failed to record events briefer than 40 or 50 msec, and small image fields that excluded the pharyngeal region. Current technology, using synchronized pulsed radiation and improved image intensifiers, permits faster camera speeds, better temporal resolution of movements, better coverage of the entire vocal tract, and smaller exposures. Data on muscle involvement are obtained from electromyographic recordings (see the section of Chapter 96 by Hirose).

Perkell (1969) gives a full set of profiles for a speaker of American English, traced from a radiographic motion film of continuous speech. Early collections of profiles such as the classical studies by Holbrook (Carmody, 1936, 1937) were traced from still pictures of sustained positions.

Linguistic Considerations

A brief word on the linguistic aspect is relevant to a fuller understanding of the role of the different parts of the vocal tract. When we speak, the tongue, lips, mandible, velum, larynx, and vocal folds are moved into different positions to generate sound or modify the spectral quality of sound. The timbre of the sound heard can be transcribed using a phonetic alphabet and square brackets, or example, [khæt] for "cat". Sound segments that have no meaning are strung together in ordered sequences to form words that do have meaning. The initial consonant sounds [f, o, s f] in the words "fin", "thin", "sin", and "shin" consist of different qualities of hissing and constitute in these examples the only audible difference between otherwise identical sounding words. Minimal segments that differentiate between meanings in this way are called *phonemes* and are denoted by slashes (in this example /f, o, s, f/). Phonemes are abstract, language-specific, distinctive units whose actual physical expression varies from context to context. For example /l/ in "coal" is voiced, resonant, and "dark", whereas in "clock" it is voiceless hissed. Such physical variants of phonemes are called *allophones*. The order of phonemes can be varied to express different meanings as in "stack", "tax", "sacked", "acts", and "cats". A fuller account of phoneme theory and the phonemic structure of English can be found in textbooks of phonetics such as that by Ladefoged (1982). Examples of contrastive spectral features by which we discriminate between the realization of phonemes are the weak hissing of /f, o/ versus the strident hissing of /s, f/; further, the spectral energy of /f/ is concentrated around 2500 to 3500 Hz, whereas for /s/ it is around 5000 to 7000 Hz. One can continue like this and list the distinguishing spectral features for each phoneme in a language (see for example Jakobson et al, 1952).

The flow rate of linguistic formation is high, about 5 syllables per second (between 10 and 20 phonemes per second). One consequence is that, while phonemes are ordered sequentially, their physical attributes are not strung along individually like the discrete letters of this page. Instead, the component maneuvers expressing successive phonemes are woven together in such a way that the acoustic cues for any single phoneme are scattered along the speech wave and that at any given moment the speech wave simultaneously contains acoustic cues for several adjacent phonemes (Fig. 95-23). This is known as coarticulation, and it has been argued that it is the unique principle of the Linguistic encoding of the speech wave (Liberman et al, 1967). In any case it is a mode of patterning that requires strict temporal coordination between individual maneuvers (Wood, 1991).

Source-Filter Theory

The air enclosed in the vocal tract is excited by vibrations generated at some sound source. The sound sources for speech are the vocal cords (glottal tone for vowels and voiced consonants and aperiodic turbulence for /h/) and constrictions in other parts of the vocal tract (aperiodic turbulence for hisslike consonants such as /f, s/ and the plosive burst after occlusive consonants such as /p/t/k/.

The periodic vibration of the vocal folds modulates the expiratory airflow, producing a train of pulses compounded of numerous simple sinusoids that appear in the voice spectrum as harmonics at integer multiplies of the fundamental frequency. The amplitudes of higher harmonics decrease by about 12 dB/octave.

At most frequencies, reflected vibrations interfere with outward-going waves, but at resonance frequencies they coincide everywhere, and standing waves are formed extending from the larynx to the lips. Consequently, all parts of the vocal tract contribute to each resonance mode. Frequency components that are close to a resonance mode are transmitted strongly through the vocal tract while other frequency components are attenuated. The frequencies of resonance are tuned by combining different maneuvers to make local adjustments to vocal tract shape. The pharynx is modified by altering the position of the back of the tongue, the degree of contraction of the pharyngeal constrictors, and the height of the larynx. English /r/ is an example of a consonant with pharyngeal narrowing to filter the glottal tone without actually creating a pharyngeal turbulence source (Fig. 95-24).

Quantal Aspects

Stevens (1972) has pointed out that the relationship between articulation and acoustic variables is not linear over the full range of a speech maneuver and that the resulting acoustic discontinuation in the speech output appear to divide the articulation of speech sounds into natural quanta. One example is the acoustic discontinuity between vowel-like sounds (constriction cross-sectional area at least 0.3 cm^2, glottal source) and hiss consonants (critical constriction about 0.2 cm^2, local turbulence source, antiresonance), and between the latter and complete occlusion. Another example is the location of the constriction for vowels. By manipulating a simple tubular model of the vocal tract, Stevens found that the output was not sensitive to small shifts of location at the hard palate, at the soft palate, or in the pharynx. Wood (1979) found by analyzing radiographic data from different languages that four such quantal locations are actually used in speech: along the hard palate for (i, e)-like vowels (English "beat", "bit, "bet"), along the soft palate for (u, U)-like vowels (English "cooed", "could"), in the upper pharynx for (o, o)-like vowels (English "caught"), and in the lower pharynx for (a, A, a)-like vowels (English "cat", "cut", "cart") (Fig. 95-25). Fig. 95-26 gives a small selection of profiles to illustrate these locations, of which two occur in the pharynx.

The musculature of the tongue is admirably situated for creating these constrictions (Fig. 95-27), especially the three posterior locations. The styloglossi and hyoglossi have a strongly directional action, and the palatoglossi and the pharyngeal constrictors have a strictly local sphincter action. No muscle can pull the tongue in the desired direction for the palatal vowels; it has to be pushed and guided into position by the genioglossi and by intrinsic muscles. All these muscles act in one way or another on the pharynx by widening or narrowing it locally. The profiles in Fig. 95-26 show that the pharynx is widest for palatal (i-e)-like vowels. For velar (u-U)-like vowels the vocal tract is narrowed at the faucial isthmus and widened in the lower pharynx. For pharyngeal (o-o)-like vowels the upper pharynx is narrow and there is slight widening near the epiglottis. For pharyngeal (a-a)-like vowels the pharynx is very narrow, especially in the lower section. Each of these vowels, like any other vowel and like any consonant, has a typical setting of the pharynx.

Perturbation Theory

The contribution of a particular speech maneuver to the spectrum depends on its relation to the nodes or antinodes of the standing waves: widening near a volume velocity node lowers the frequency of that resonance, whereas narrowing raises it (vice versa near an antinode).

Fig. 95-28 shows the approximate locations of the nodes and antinodes for four resonance modes. For each resonance there are at least a node in the larynx and an antinode at the lips. There are an additional node and antinode within the vocal tract for each higher resonance mode. Pharyngeal narrowing will raise F1 (proximity to the laryngeal node) and lower F2 (pharyngeal antinode). Narrowing in the upper pharynx raises F3 (uvular node), while narrowing the lower pharynx lowers F3 (pharyngeal antinode). Widening has the opposite effect. For F4, the nodes and antinodes are so close that maneuvers cannot be selective. On the other hand, lowering the larynx decouples F4 from the pharynx, which is useful in singing to bring F4 closer to F3 to produce the "song formant" (Sundberg, 1970, 1977).

How sensitive is a resonance mode at any particular node or antinode? This is related to the energy distributions: the sensitivity of a resonance frequency to a local maneuver is proportional to the local difference between the kinetic and potential energy of that resonance mode. An example is given in Fig. 95-29. F2 of an (i)-like configuration is less sensitive to further perturbations in the anterior part of the vocal tract and more sensitive in the pharynx. Fant has emphasized how this contradicts the obsolete traditional view that F2 represents a "mouth resonance". The corresponding sensitivity diagrams for the first four resonance modes of a selection of vowels illustrated in Fig. 95-30 demonstrate the sensitivity of resonance modes to maneuvers in the pharynx.

For a quantitative appreciation of the effect of a specific maneuver on the output, the actual resonance frequencies need to be found. The spectral consequences of systematic modifications to vocal tract shape have been presented by Stevens and House (1955), who used a simple tubular model in which the size of the orifice and the size and location of a parabolic internal narrowing were varied, and by Lindblom and Sundberg (1971), who manipulated the positions of the lips, mandible, tongue, and larynx on a vocal tract profile. Wood (1992) investigated the contribution of pharyngeal maneuvers to the spectral contrast between vowel phonemes in vowel pairs such as "beat-bit", "cooed-could", "caught-cot" (or "coat-caught" depending on the dialect), and "cart-cut" (Figs. 95-26, 95-31 to 95-33). The results in Fig. 95-33 show the contribution of the pharyngeal modification compared with the contributions of other maneuvers involved in each case. In pairs like this, the sound contrast is achieved largely by a small adjustment in the width of the lower pharynx made by varying the activity of the posterior genioglossi, accompanied by an associated difference in the degree of constriction and by differences in lip activity (Raphael and Bell-Berti, 1975). Such contrasts are found in many other languages. Similar pharyngeal behavior in vowel harmony in certain African languages has attracted attention (Jacobson, 1980; Painter, 1973; Vago, 1980).

Motor Control

The rapid flow rate of speech (10 to 20 phonems per second) and the interwoven coding pattern of speech articulation require precision in tuning the resonances of the vocal tract and in coordinating the timing of each individual maneuver and of each sound source. This precision is ensured by close motor control and supported by the differentiated afferent feedback, fast reflex arcs, and proprioception provided by the interconnections and central projections of the cranial nerves (Bowman, 1971; Dubner et al, 1978; Fucci and Petrosino, 1981; Grillner et al, 1982; Kelso, 1982).

The critical aperture dimension for turbulent sound sources is an example of this precision. Turbulence arises in a constriction with an aperture of about 0.2 cm^2 (a cross distance of 1.5 to 2 mm). Thus a tolerance of only about 1 mm separates turbulence generation from complete occlusion and from turbulence-free vowel-like sounds. The afferent information on the pattern of contact at narrow constrictions comprises tactile sensations from the surface of the tongue and the mucosa of the constricted part of the vocal tract (lips, gums, palate, velum, or pharynx). A similar tolerance was demonstrated in model experiments by Wood (1982) for the tongue positions of palatal vowels. In this case the configuration is maintained by balancing tongue elevation against mandibular depression. Lindblom et al (1979) demonstrated that the synergy is extremely efficient; subjects with bite blocks placed between their teeth are able to hit the correct vocal tract configuration on the first glottic pulse of the vowel. In addition to the tactile sensation from the edges of the tongue, the gums, and the palate, the control of this maneuver also receives proprioceptive information from the mandibular joint and from the spindles of the tongue (lingual maneuvers), masseter muscle (mandibular elevation), and anterior belly of the digastric muscle (mandibular depression). Much of this efferent feedback flows near the pharynx.

Summary

This section has outlined the basic theory of speech production, with examples of the various ways in which the pharynx participates in speech. First, as a cavity of the vocal tract the pharynx takes an active part in the generation and filtering of sound. Second, the muscles of the pharyngeal region (including the extrinsic muscles of the tongue) have both a local role in shaping the pharyngeal cavity and a more general role in positioning the tongue in relation to the anterior part of the vocal tract. Finally, most of the efferent and afferent nerves that mediate the fine control of speech maneuvers throughout the vocal tract pass in the vicinity of the pharyngeal region. A peripheral lesion may thus leave the patient with unfamiliar cavity proportions, reduced oral agility, or impaired precision and coordination. Since radical surgery is often a lifesaving necessity, a weakening or loss of some function may be an inevitable price. The question remains: can expected impairment be ameliorated by planning the intervention differently?