Chapter 96: Neurophysiology

Colin Painter

Voice Evaluation

Margaret L. Stoicheff

After laryngoscopic examination of the patient with a voice disorder, the laryngologist may make a referral to the speech pathologist for a voice evaluation. This evaluation involves a detailed case history; measuring various aspects of voice using clinical instrumentation when available (Baken, 1987); observing visually and auditorily how the patient is producing voice; and testing the patient's ability to alter voice in response to various manipulations/instructions/voice techniques (that is, therapy). The results of this voice evaluation combined with the laryngeal findings enable the speech pathologist to assess the contribution of physical, psychologic, and production factors on voice and to determine how amenable the disordered voice may be to direct work. The voice evaluation provides important insights into the individual's motivation and readiness to improve voice. It may reveal that the patient's voice problem is an expression of a deep-rooted emotional problem requiring referral for psychotherapy (Aronson, 1990) or that the voice problem may be the first symptom of a motor speech problem, requiring referral to a neurologist (Aronson, 1971; Darley et al, 1975).

Case History Information

A detailed case history provides indispensable information for making management decisions. There are a number of recent voice texts that cover history taking in some detail (Aronson, 1990; Boone and McFarlane, 1988; Colton et al, 1990).

Complaint

The interview typically begins with the patient's own description of the voice problem. This includes the patient's awareness of the specifics of the vocal problem and what he or she believes to be its basis. It is essential to ask about associated symptoms; these may shed some light on the nature of the problem or may provide a useful index for assessing change in response to treatment. In addition, the patient's appraisal of the effects of the voice disorder on his or her life provides an indication of how realistically the patient is viewing the problem and how effectively the patient is coping.

Onset and development of problem

Questions to ask include the following: when the problem started, whether onset was sudden or gradual, and whether there have been changes in the problem over time. These may point toward an organic or psychogenic etiology. The variability of the problem during the day or in response to certain activities can provide further diagnostic insights. An exploration of why assistance is being sought at this time as well as previous attempts to obtain help may aid in assessing the motivation of the patient to make voice changes.

Voice usages

The speech pathologist carefully explores the ways in which the patient uses voice on the job as well as in social, athletic, and community activities. The environments in which this speaking occurs (dust, fumes, lack of humidity, poor acoustics, inadequate amplification, noise level) may be causally related to the presenting problem or may be maintaining it. The timing of vocal usages may be critical: for example, excessive speaking or singing in the evening after a vocally demanding day, speaking or singing before groups when the individual has laryngitis or is unwell, or excessive postperformance talking (Colton et al, 1990). Personal factors requiring investigation include such areas as the professional voice training that the singer has had; the range of voice required for a specific role in a play or musical; the presence of a hard-of-hearing relative in the home; and the vocal image that the individual is attempting to project on the job.

Medical history

Moore (1971) has stressed the importance of good *physical health* to optimal voice functioning. Aronson (1990) has highlighted the importance of good *emotional health*, stating that "the extrinsic and intrinsic laryngeal muscles are exquisitely sensitive to emotional stress, and their hypercontraction is the common denominator behind the dysphonia and aphonia in virtually all psychogenic voice disorders".

Obviously, the voice history must focus on health status, present and past. It is important to know whether the patient has any respiratory problems, allergy-related problems, neurologic disorder, anemia, thyroid dysfunction, hearing loss, or any other current health problem. It is also necessary to inquire into past medical history such as hospitalizations and surgical episodes, particularly those involving the head and neck. All medications and drugs that are being used need to be checked to assess their effects on laryngeal mucosa and laryngeal motor control (Colton et al, 1990). Smoking habits also need to be considered because of their known effects on the larynx and voice.

Voice Testing

The speech pathologist systematically investigate how the individual is producing voice as well as the voice itself.

Respiration for speech

The way in which the person is breathing for speaking is routinely assessed. This involves scrutinizing speaking behaviors to determine whether the individual is operating within the most efficient lung volumes for speech (Hixon, 1987). Observations are made on the length of the breath groups for speech, the frequency of glottal fry phonations, the audibility of inhalations, and the amount of audible/visible effort being expended for breathing. Attention is also directed toward the timing of inhalations for speaking and whether these intakes are predominantly oral (a requirement for speaking) or nasal. If the air volumes used for speaking are frequently in the less accessible expiratory reserve volume range, the work involved in voice production is increased and may have a significant negative impact on voice.

A simple clinical test to measure respiration potential for speech as well as adequacy of laryngeal valving for phonation is the maximum function test known as the s/z ratio (Eckel and Boone, 1981). Following demonstration by the clinician, the patient is instructed to inhale as much air as possible through the mouth and then to sustain the voiceless /s/ sound for as long as possible. Production of /s/ (a hissing sound) requires no laryngeal valving. This task is repeated three times or until the clinician is convinced that the maximum duration possible has been achieved. The same procedure is followed for eliciting a maximally sustained /z/ sound. The /z/ sound requires the same respiratory control as the /s/ with the added component of laryngeal valving or voicing. If the values of both sounds are abnormally short in terms of values reported in the literatur, a problem in respiratory control would be expected. However, if the /s/ value falls within normal limits but the /z/ value is shorter, then the size of the s/z ratio needs to be considered. Eckel and Boone (1981) have reported that the s/z ratio approximates 1.00 in subjects without laryngeal pathologic findings and that it exceeds 1.40 in 95% of patients with laryngeal pathologic findings. Therefore the magnitude of this ratio can be used to estimate changes in laryngeal pathologic conditions with therapy. A maximally sustained vowel, usually (ah), provides further information on laryngeal function.

If the voice clinician has access to airflow equipment, such as the Phonatory Function Analyzer, it will also be possible to look at the amount of airflow through the glottis as the patient phonates. This will in turn provide some information on normalcy of laryngeal valving for voicing.

Control of frequency (pitch)

The control of pitch is assessed in a number of ways that interrelate to provide a picture of the patient's ability to lengthen and to shorten the vocal cords.

The speaking pitch, or the *fundamental frequency,* of the voice is measured using instrumentation such as the Visi-Pitch or Vocal II. In the absence of instrumentation, laborious procedures involving tape recorings and matching with a pitch pipe may be used. The obtained fundamental frequency value is compared to normative data available for men and women at various ages. The fundamental frequency of the voice, however, must be viewed in context for the individual patient. For example, if there is webbing of the vocal folds, the fundamental frequency will be higher than the normative values because the effective vibrating length of the cords is shorter. In this case the speech pathologist could measure fundamental frequency preoperatively and postoperatively to document the effects of surgical intervention. The occurrence of a high fundamental frequency in the presence of vocal folds of adult length and normal motility indicates the need for direct work on lowering pitch. Thus knowledge of the laryngoscopic findings is crucial in interpreting the results of voice testing.

Another aspect of the frequency that is measured is the *phonational frequency range (PFR)* of the patient. The patient is asked to produce an ascending and descending series of phonations, usually using the equal-tempered musical scale. The task is to determine the highest and lowest frequency the person can phonate. Using a matching procedure (piano tones or pure tones) or the Visi-Pitch th patient starts at a comfortable pitch and sings down to the lowest possible tone. Then the patient starts at a comfortable pitch and sings up the scale to the highest possible tone (including falsetto). The range between the highest and lowest sung tones, the PFR, should include a minimum of two octaves, or 24 semitones, according to Hollien (1977), who has provided much of the normative data. Although it may take some practice to elicit the PFR from a patient, this does provide useful information. It gives a good indication of the normalcy of function of the vocal folds. It is not unusual to find that the patient with a severe voice disorder in speaking can produce a three-octave range or greater, substantiating negative laryngoscopic findings. Conversely, the inability of a person to produce a range of at least two octaves will create some serious question of the normalcy of function of the vocal folds. During this testing, the speech pathologist makes note of the quality and of the ease of production of the tones produced throughout the range. It is important to document whether the patient with a severe dysphonia in speaking becomes progressively less dysphonic as he or she attempts to produce these tones or whether the patient experiences similar voice production difficulties in this situation as in speaking.

Flexibility of pitch in the voice is important to efficient and effective use of the voice. This flexibility may be restricted or excessive for a number of reasons. Using the inflection patterns found in the English language, the speech pathologist tests the patient's ability to make such rapid pitch changes in speech.

Control of Intensity

Another dimension of voice that is critical to assess is intensity, or loudness. The intensity of the *conversational voice* in decibels can be measurd using a sound level meter. Researchers (Timcke et al, 1959) using high-speech motion picture photography of the larynx have shown that in loud productions the vocal folds are in contact for a greater part of the vibratory cycle and that the folds snap back to the adducted position after their opening phase with more force than in less intense productions. Use of an habitually loud voice (or an elevated intensity level) therefore is potentially irritating to vocal fold tissues. On the other hand, use of an habitually soft voice indicates less-than-optimal adduction of the vocal folds for voicing purposes.

The *range of intensities* the patient can produce is tested to assess the ability to approximate the vocal folds sufficiently to build up the necessary subglottal pressures for such a range. The patient is asked to produce a word such as *no* at very soft, soft, moderately loud, and very loud levels. The speech pathologist not only notes the range that can be produced but also observes how such changes are effected. Does the patient use various lung volumes to produce the changes? Does he or she make use of sufficient mouth opening to effect the changes?

Voice quality

The area of voice quality is the area in which there would be some diagreement on the particular terminology employed to describe voice. However, there is uniformity on how speech pathologists test a patient's ability to change the quality of the sounds produced.

Many speech pathologists use adjectives such as *breathy, hoarse, harsh, rough,* and *strained* or *squeezed* to describe voice quality and provide some measure of the degree of severity of the deviant quality using a numeric scale. Whatever the system, there can be good communication between the laryngologist an the speech pathologist when the two agree on the perceptual labels use and are consistent in their application. For those speech pathologists fortunate enough to have spectrographic equipment available to them, sound spectrograms of the voice can be made to provide a visual record of the quality deviation. Jitter data may also be available from special programs for the Visi-Pitch. In all instances the speech pathologist will make tape recorings that can be used at a later date for perceptual comparisons of changes in quality. In many settings the speech pathologists will make these evaluations more objctive by having other professions make ratings without knowlege of the patient or the management.

To determine the variability of the quality deviation, the speech pathologist will systematically have the patient vary pitch and loudness an note the changes in voice quality. It is sometimes possible to reduce or eliminate rough quality by a slight elevation of pitch. Breathy quality may be improved by an increase in pitch or loudness or both. Of course, any such changes must lie within the acceptable range of productions for that patient to be consiere as possible therapy approaches.

Positive variations in quality may also be effected by changes in the muscular set for phonation. Through experimentation with various muscle loosening techniques such as chewing, yawning, and sighing, a different (more optimal) set of oral, pharyngeal, and laryngeal musculature can be facilitate, at least temporarily. This permits observation of the way in which quality changes as a consequence. Other aspects of production, such as phonation volume, are also varie in order to note the effects on the voice quality.

The conscious manipulation of voice along a number of parameters will result in a more dramatic change in quality in some patients than in others. It would appear that this experimentation forces some patients out of a habitual laryngeal or phonatory set that is counterproductive. This process provies the patient with some insight into the nature of the problem and may demonstrate the possibility of making changes. In many instances such experimentation provides the starting point for the beginning of therapy.

Abusive laryngeal behaviors

During the assessment, the speech pathologist pays particular attention to the frequency and force of throat-clearing behaviors, coughing, and glottal attacks (coups de glotte). Many patients will clear the throat or cough before initiating phonation and are unaware of these behaviors or think that they must do them in order to phonate. Because of these behaviors involve massive, forceful contact of the vocal fols, their frequency will determine their significance for the presenting voice symptomatology.

Glottal attacks in speech are considered in terms of their frequency and force. Normal speech is not free of glottal attacks in certain contexts. However, the force and frequency of glottal attacks need to be assessed in terms of the role these may play in maintaining the patient's difficulties.

Associated factors

Although speech pathologist concentrates on the voice in this evaluation, the voice problem must be considered within the framework of speech in general. Thus adequacy of resonance, rate of speech, articulation, and language are informally gauge by the speech pathologist to determine if there are other communication problems present that might be associated with the voice disorder. A peripheral oral examination is performed to note the presence of any contributory structural or motor deviations. The acuity of hearing is screened to see whether audiologic testing is required. It is important to be awar of a hearing problem, both for evaluation of the voice problem and for its possible implications for management.

Throughout this initial session the speech pathologist attempts to develop an unerstanding in the patient that voice is the product of many muscular events and that it does not just "happen". This is done through explaining what is being tested, through diagrams or models of the larynx, and by answering questions as they arise. The voice evaluation is an important part of the therapy process and is where therapy begins.

At the conclusion of this voice evaluation, the speech pathologist will provide the otolaryngologist with a detailed report on the ways in which the voice is disordered, life-style factors and vocal behaviors that may be related to the disorder, recommendation for management that may or may not inclue voice therapy, and prognosis.

Videolaryngoscopy and Laryngeal Photography

Eiji Yanagisawa, Roy D. Carlson

This section describes videographic and photographic documentation of the larynx in the office.

Videolaryngoscopy (Videography of Larynx)

Videolaryngoscopy - a videographic documentation of endoscopic laryngeal examination with simultaneous voice recording using a video camera - is a most useful office procedure for the clinical practice of laryngology. It allows excellent visualization and documentation of the physiologic functions and pathologic conditions of the larynx. Videolaryngoscopy can be accomplished using either a flexible fiberscope or a rigid telescope (Fig. 96-1).

Since the introduction of flexible fiberscopic laryngoscopy by Sawashima and Hirose in 1968, the flexible fiberscope has been widely used for documenting various conditions of the larynx, including evaluation of postlaryngectomy speech, spastic dysphonia, stuttering, and even difficult endotracheal intubation. The ease an effectiveness of photo and video documentation by the flexible fiberscope have been emphasized (Selin, 1983, 1984; Yamashita, 1988; Yanagisawa et al, 1983, 1986).

One of the other significant advances in laryngeal examination has been the development of the rod lens telescope by Hopkins and its applications to laryngeal documentation by Ward et al and others (Berci, 1974; Konrad et al, 1981; Ward et al, 1974;

Yanagisawa et al, 1981, 1983, 1984). The use of the telescope for video documentation of the larynx has been advocated by many authors. Yanagisawa et al (1983, 1984) felt telescopic videography is superior to fiberscopic videography in precise documentation of structural changes of the larynx. A simple and inexpensive method of videolaryngoscopy using a telescope and a home video system camera was described by Yanagisawa et al in 1981.

More recently, videolaryngoscopy has been combined with stroboscopy to observe vocal cord vibrations in health and disease (Bless et al, 1987; Sataloff et al, 1988; Yanagisawa et al, 1987).

Equipment

Flexible fiberscope

The commonly used flexible fiberscopes are the Olympus ENF-P2 or ENF-P3, Machida ENT 4L or 3L, and Pentax FNL 10S or 15S. Since the optical end of each fiberscope is different, each fiberscope requires a specific video camera adapter. The Olympus ENF-P2/P3 fiberscope is our choice. The outer diameter of the insertion tube of this fiberscope is 3.4 mm. The simple thumb movement control of tip bending in the Olympus fiberscopes appears to be an obvious advantage over the design with the know on one side of the control head, which calls for two-handed manipulation. This is particularly true when the scope is used with a video camera.

Rigid right-angled telescope

There are several excellent right-angled telescopes especially designed for laryngeal examination. They are (1) Berci-Ward Karl Storz 3702D, (2) Karl Storx 8704D, (3) Karl Storz 8706CL, (4) Wolf Stuckrad, and (5) Nagashima SFT-1. The prism angle of the Nagashima SFT-1 (our choice) and Karl Storz 8706CL telescopes is 70 degrees, permitting excellent visualization of anterior and posterior commissures of the larynx, and is particularly useful for laryngostroboscopy. These telescopes are connected to the front of a home video camera (Fig. 96-2) via the Karl Storz quick connect adapter (Fig. 96-4, left) and to a miniature CCD camera (Fig. 96-3) via a specific adapter (Fig. 96-4, middle and right). We use primarily the Nagashima SFT-1 rigid, right-angled telescope (see Fig. 96-1) with the "Olympus type" eyepiece. The Nagashima laryngoscope is a single integrated unit composed of a telescope and a light transmission system. It measures 24.5 cm in length and 9 mm in outer diameter. The angle of the prism at the tip of the laryngoscope is 70 degrees. The rectangular outlets on each side of the lens offer well-balanced illumination for the viewing field.

Video camera

Many excellent compact video cameras are particularly designed for medical use, such as (1) Karl Storz Mini 9000 (17 Lux), 9050B, S9080E CCD (C-mount); (2) Olympus OTV-F2; (3) Toshiba (Nagashima) CCD (10 Lux); (4) Elmo EC-202 CCD (10 Lux) (Fig. 96-3); (5) Elmo EC-102 CCD (Olympus) (15 Lux) (Fig. 96-3); (6) JEDMED Micro CCD (8 Lux); (7) Storz ENT 62 CCD (7 Lux); (8) Olympus MC-6 CCD; and (9) Wolf CCD (10 Lux). These cameras are equipped with C-mount couplers and require special adapters for the endoscopes (Fig. 96-4). The Karl Storz Supercam 9060B CCD camera (7 Lux) has a built-in beam splitter (no C-mount coupler) (Fig. 96-3). More recently, a high-resolution three-chip CCD mini camera (Sony DXC-750MD: 750 TV lines, 20 Lux; approximately $16.000) has been introduced. Although small video cameras are compact and convenient, their cost is often prohibitive to many otolaryngologists. We advocate the use of reasonably priced home video system cameras for endoscopic videolaryngoscopy and have obtained excellent results. Home video cameras used are (1) JVC GX N8U (10 Lux); (2) Olympus Movie 8 VX801-62 (7 Lux), and (3) Ricoh R620 CCD (4 Lux) (see Fig. 96-2). These home video cameras have a macrozooming device, but the lens is not detachable. With the use of a Karl Storz quick connect adapter, however, a fiberscope can be attached to the front of the lens.

Light source (Fig. 96-5)

The Xenon cold light fountain, which produces approximately 6000 K, is recommended for both fiberscope and telescopic video documentation. However, recently developed video cameras are so light sensitive that standard light sources, such as the Olympus ILK-3, Machida LH-150, Karl Storz 481-C, or Pilling Luminator 2X, may suffice for telescopic videolaryngoscopy. For fiberscopic documentation, however, the more powerful light sources, such as the Karl Storz Xenon light source 487C, Karl Storz metal halide arc (484C), Olympus CLE-F10, or Olympus CLV-F10, are still necessary to produce consistently high-quality video images. Karl Storz recently produced a smaller and "quieter" Xenon light source 615.

Video camera adapter (see Fig. 96-4)

A home video camera can be used with an appropriate adapter attached to the front of the video camera lens (see Fig. 96-4). A miniature CCD camera is connected to the endoscope using a special C-mount adapter. Some have a built-in adapter such as the Karl Storz Supercam 9060B.

Video recorder and monitor

Technically, any 3/4 inch, 1/2 inch, or 8 mm video recorder can be used. When a camcorder is used, documentation can be made onto the camcorder. However, recording can also be made onto a separate video recorder (Fig. 96-6).

Technique

Fiberscopic videolaryngoscopy (Fig. 96-7)

The Olympus ENF-P2/P3 or the fiberscope of your choice is connected to a home video camera or to a miniature CCD camera with a C-mount coupler using an appropriate adapter. The patient sits facing the examiner, who may be either sitting or standing. Both nasal cavities are sprayed first with 3% ephedrine to shrink the turbinates and then with 2% tetracaine (Pontocaine) or 4% lidocaine (Xylocaine) spray. The tip of the flexible fiberscope is lightly lubricated and inserted into the nose gently. If the patient complains of pain, the procedure is stopped. Pain usually means that the tip of the fiberscope is hitting the septal spur or abnormally positioned or enlarged turbinates. The nose is reexamined, and an attempt is made to pass the scope either below or above the spur or to use the opposite side. It may be necessary to respray the nose. While the examiner is observing through the camera viewfinder or on the monitor, the fiberscope is gradually advanced to the nasopharynx. The tip is bent downward, and the scope is passed to the level of the epiglottis where the entire vocal cord can be visualized.

Mucus in the hypopharynx may cause clouding of the lens of the fiberscope. This can be easily corrected by having the patient swallow several times, which helps to wipe out the small distant lens of the fiberscope. The larynx is centered on the screen, and the patient is ready for videotaping. Some examples of fiberscopic images are shown in Fig. 96-8.

Telescopic videolaryngoscopy (Fig. 96-9)

The patient and examiner sit facing each other. The soft palate and the posterior half of the tongue are topically anesthetized with 2% tetracaine or 4% lidocaine. The tip of the telescope is dipped in warm water and dried to prevent fogging of the lens during the examination. The video camera and illuminator are turned on. The xenon illuminator is set in the medium filming mode. The examiner holds the video camera in one hand and inserts the telescope into the mouth while holding the patient's tongue with the other hand. To prevent the clouding of the telescope lens, every effort should be made to ensure that the tip of the telescope does not touch the base of the tongue or the uvula. This can be accomplished by having the patient say (ee) as the scope is inserted. When the desired image appears on the monitor screen, the video recording is begun. The laryngeal image is centered on the screen using the viewfinder of the camera or the TV monitor. The light intensity is adjusted according to the video image desired. Some examples of telescopic images are shown in Fig. 96-10.

Comments

Fiberscopic videolaryngoscopy has great value in the diagnosis of functional voice disorders and in the evaluation of laryngeal function, as well as the diagnosis and documentation of pathologic conditions (see Fig. 96-8). Respiration, phonation, glottal effort, closure, and swallowing can be demonstrated effectively with the fiberscope. It is very useful for voice and speech evaluation, particularly for functional disorders such as spastic dysphonia, psychogenic dysphonia, and stuttering (Conture et al, 1977; Parnes et al, 1978; Williams et al, 1975). It is also very useful for singers to study and improve their own methods of singing.

One of the most significant advantages of fiberscopic videography is the ease of examination. Fiberscopic videography is the procedure of choice for children or hypersensitive adults whose larynx cannot be examined either by mirror or telescope.

Some of the disadvantages of fiberscopic videography are (1) it provides less clear and more distorted pictures of the larynx; (2) it requires a very powerful light source (xenon) because of the small size of the fiberscope and the long distance between the larynx and the camera; and (3) it may not demonstrate minor mucosal changes and early lesions of the vocal cord.

Telescopic videography provides much clearer, sharper, brighter, and larger images of the larynx (see Fig. 96-10). For documentation of precise anatomic or structural changes of the larynx, we find telescopic videography of the larynx to be far superior. The most significant advantages of telescopic videography are its clear video image and high resolution. Although from a technical viewpoint telescopic videography is not ideal for evaluation of speech and singing, the vibration of the vocal folds on phonation and singing is much more clearly shown using a telescope than by fiberscopic videography (Yanagisawa, 1984; Yanagisawa et al, 1983).

Some of the disadvantages of telescopic videography are (1) children and some adults with hypersensitive gag reflexes may not be able to tolerate the procedure; (2) fogging of the lens during the procedure occurs more commonly; (3) normal speech and singing voice are distorted because the tongue is held and pulled forward; and (4) the need for local anesthesia may alter normal laryngeal function, especially phonation and swallowing.

Both fiberscopic videolaryngoscopy and telescopic videolaryngoscopy play important roles for the evaluation and management of disorders of the larynx and are of great value for patient care, teaching, voice analysis, and documentation of various laryngeal disorders.

One of the most important advantages of videolaryngoscopy is the ease of simultaneous voice recording. Voice can be recorded simply by plugging the audio output connector of the home video camera equipped with a built-in microphone to the cassette recorder. For high-fidelity voice recording, the use of a separate external microphone directly connected to the cassette recorder is recommended. Simultaneous audio and video documentation is particularly valuable for voice analysis of the patient before and after treatment. The surgeon can document a progressive improvement or regression of the laryngeal disorders of the patient.

The taped image can be replayed and examined repeatedly. This will eliminate the need for multiple and prolonged examination of the patient. More critical evaluation of the laryngeal image can be done by freezing its motion for a frame-by-frame examination using a videotape recorder equipped with a pause and slow motion capability. This allows the physician to detect a small, obscure lesion that may be unrecognized at the time of the initial mirror examination.

For successful videography, the cooperation of the patient is of paramount importance. This is particularly true with telescopic videolaryngoscopy. To relieve the anxiety of the patient, it is important that the patient be informed of all the steps involved, including the effect of topical anesthesia ("feeling of choking", "difficulty in swallowing", and so forth). The patient should be told that such an effect is only temporary. It should be noted that inadequate topical anesthesia may be the main cause of failure in a patient with hypersensitive gag reflex.

Transfer of video images to prints/slides

Videolaryngoscopic images can be instantaneously transferred to prints using either a black-and-white or a color video printer. These video printouts are of great value for permanent record, patient counselling, and comparison of laryngeal conditions before and after treatment. Black-and-white video printers available include (1) the Sony video-printer UP-811 (Fig. 96-11) and (2) the Mitsubishi P-60U. The relatively new Sony UP-5000 color video printer is costly (approximately $7.000) but produces superb color prints of videolaryngoscopic images of high resolution in 1 minute. The print can be of either single or multiple images. Affordable color video printers such as the Sony CVP G500 (approximately $1500) are now available.

A color TV image can be photographed inexpensively. A color slide can be obtained by photographing the TV image using a single lens reflex camera with a shutter speed of one-thirtieth of a second or slower. With the Kodak CC40R or Tiffen CC40 filter with ASA 400 film, an excellent photograph can be obtained.

Still Laryngeal Photography

There are many different methods of still laryngeal photography. Here, fiberscopic photography and telescopic photography, which can be accomplished in the office, will be described.

Fiberscopic photography

Photographic documentation of the laryngeal images using the fiberscope is difficult because of the small diameter of the fiberscope and the need of a powerful light source. Selkin (1983, 1984) published excellent papers on this subject.

Equipment (Fig. 96-12)

The equipment used in fiberscopic photography includes (1) the Olympus ENF-P2 or P3 flexible fiberscope; (2) the Karl Storz Xenon cold light fountain (487C, 615) or equivalent xenon light source; (3) the Olympus SMR endoscopic coupler (Fig. 96-12); (4) the Olympus OM-2 35 mm single lens reflex camera with automatic winder an clear glass focusing screen I-9 (Fig. 96-12); (5) the 2X teleconverter (Vivitar); and (6) Kodak Ectachrome 400 ASA Daylight film.

Technique (Fig. 96-13)

The nose is first examined and sprayed with 3% ephedrine and 2% tetracaine or 4% lidocaine spray. The side of the nose without septal spur and with wider air passages is chosen. The fiberscope is advanced to the epiglottis just as for videolaryngoscopy. The laryngeal image is centered in the viewfinder of the camera, and the photographs are taken. The larynx is first examined and photographed on respiration (at rest) and phonation (while saying (e)).

The Olympus fiberscope is attached to the 100 mm Zuiko lens on the Olympus OM-2 SLR camera via the quick-connect adapter (Machida adapter for Machida fiberscope; Fig. 96-14). An alternative method is to attach the Olympus fiberscope to the camera using the Olympus SMR coupler with 2X extener. The lens is set at infinity and the aperture wide open (f2.8). The camera is set at automatic mode. The exposure is bracketed on the top of the camera. The picture is taken at 0, -1, and -2. Because of the small size of the image on the

35 mm film screen, the automatic camera has a tendency to overexpose the laryngeal image. If the -2 setting gives overexposure, then the ASA indicator is set at 800 or 1000. The fiberoptic cable of the Olympus ENF-P fiberscope is connected to the xenon light source. The xenon light source is set at the highest filming mode (to its maximum) during the exposure.

Comments

The quality of the photography of the larynx obtained with this technique is not always predictable and not as high as that obtained through the telescope (Fig. 96-15). Yet, when the exposure is bracketed and the correct exposure for the setup is established, laryngeal pictures of acceptable quality can be obtained in about 50% of the cases.

Telescopic still photography

The right-angled telescopes used for videography can be used for this telescopic photography. Our preferred method using the Nagashima SFT-1 right-angled telescope will be described (Yanagisawa, 1984.

Equipment

The equipment required for still photography of the larynx with the Nagashima system is (1) the Nagashima rigid laryngoscope, right-angled SFT-1 (see Fig. 96-1); (2) an adapter (attachment lens, f100 mm) (Nagashima); (3) the Olympus OM-2 SLR with automatic winder and I-9 focusing screen; (4) a xenon illuminator (Karl Storz 487C); and (5) Kodak Ectachrome ASA 400 film.

Technique (Fig. 96-16)

The telescope is connected to the Olympus OM-2 SLR camera via the 100 mm attachment adapter for the old Nagashima SFT-1 with the "Machida-type" eyepieve. The fiberoptic light cable of the telescope is then connected to the Xenon light source. Instead of the f100 mm Nagashima adapter, we now use the Karl Storz quick connect adapter attached to the front of the 100 mm Zuiko lens. To this Karl Storz quick adapterm, the new Nagashima SFT-1 telescope with the "Olympus-type" eyepiece is attached. The camera is set at automatic mode, and the ASA dial is set according to the ASA rating of the film used. In this technique, the patient and the examiner sit face to face. The pharynx is topically anesthetized using 2% tetracaine or 4% lidocaine solution. The examiner holds the tongue of the patient as he or she inserts the telescope into the pharynx. When the desired image is centered via the viewfinder of the camera, the output of the xenon light source is turned to its near maximum. A series of photographs are then taken in rapid succession both on phonation and respiration, utilizing the automatic winder in a continuous mode. Excellent pictures are oftenobtained during deep respiration immediately following phonation.

Comments

Satisfactory results can be obtained in over 80% of the laryngeal pictures taken with this system (Fig. 96-17). The final size of the laryngeal picture taken with the Nagashima system using a 100 mm attachment lens is too small to be shown to a large group. Rather than enlarging the original slides for presentation, we use a 2X teleconverter between the attachment lens and the camera or the 100 mm Zuiko lens with Karl Storz quick connect adapter. In this way, the full-screen size view of the larynx ready for projection can be obtained on the finished slide.

Telescopic photography with the Nagashima right-angled laryngoscope is an easy way to photograph the larynx in the office. It produces satisfactory color photographs of the larynx suitable for teaching, permanent records, and preoperative and postoperative documentation of various laryngeal disorders (Fig. 96-17).

Xeroradiographic, Laryngoscopic, Spectrographic, and Glottographic Indicators of Laryngeal Function

Colin Painter

Phoneticians find it convenient, when undertaking the kind of work that is often called "speech science", to talk about the speech communication process in terms of a chain of events that links the speaker to the listener. Neuroscience tells us so little about the language centers of the brain that our best information comes from the grammars (that is, theoretical models) of the linguist. However, from a medical standpoint, much more is known about speech production, about the gestures and configurations of the organs of speech. Disturbances to the air beyond the lips of a speaker, part of the field of study known as acoustic physics, have been particularly well described since 1948. The ear of the listener has been known much longer but is considered by the speech scientist as little more than a transducer, reception being less intriguing than perception, that area of experimental psychology that seeks to determine how the nervous system uses incoming speech signals.

Linguistics, physiology, physics, acoustics, reception, perception - the same message transforming itself through different media - these aspects of the speech chain are so different that the means used to obtain data at various levels will clearly have to be different also. Further, information obtained by any given instrumental technique, although appropriate to the immediate task, will have its limitations. It is obvious that multiple data channels are advisable so that each may supplement and reinforce the others.

This section demonstrates how, for three basic voice qualities, multiple data - perceptual, acoustic, physiologic, and physical - may together facilitate a fairly complete picture of the larynx in action.

Xeroradiography

X-ray photography of the larynx is discussed in Chapter 97. Its primary advantage is the ability to display structures that are otherwise not open to imaging. Its main disadvantage is the risk associated with radiation dosage. However, in spite of this, xeroradiography can be particularly useful. By using it with a copper screen to shield out hard tissue such as the laryngeal cartilages, beautifully clear views of the soft tissue of the speech tract can be obtained. The dosage for five images is equivalent to that for a set of dental radiographs, which is not excessive.

Figs. 96-18 to 96-20 are sagittal views of a speaker uttering the vowel "ee", with normal, breathy, and constricted voice qualities. It should be noted that the tongue is to the same degree high and forward in the mouth in all three figures. The differences lie in the oropharynx and laryngopharynx. Fig. 96-19, with breathy voice, shows a particularly wide upper and lower pharynx, a wide laryngeal vestibule, and a markedly open laryngeal ventricle. In contrast, Fig. 96-20, with tense voice, shows a very constricted upper and lower pharynx, a narrow entry to the larynx, and a barely visible ventricle. Not all occurrences of breathy voice have similar configurations, although most cases of tense voice do. The relationship between the vocal folds, hyoid bone, mandible, and vertebrae is easily seen. The differences between these three figures are quite dramatic, especially when considering the fact that the oral settings are almost identical.

Laryngeal Photography

The following part of this section discusses available instrumentation and techniques. The advantages are apparent: permanent documentation for medical and legal purposes, the ability to monitor changes over time, the benefits of discussion with the patient, the size and clarity of the picture, and the potential for measurements. The only real disadvantage is the cost. Not discussed in the following pages are the kinds of measurements that can be made from photographs of the larynx. Usually measurements of glottal length and glottal width are the only ones that are made; if that is all one is looking for, that is all one will see. I believe that valuable information about laryngeal function can be obtained from measurements of the laryngeal vestibule. The distance between the cuneiform cartilages (as opposed to the arytenoids), the aryepiglottic fold length, the epiglottis position, the forward movement of the cuneiform and arytenoid cartilages, and the distance between the false folds - all these are important indicators of how the larynx is functioning (Painter, 1986, 1991a; Painter et al, 1991).

Plate 16 shows views while the vowel (e) (as in bed) is uttered with normal (modal), breathy, and constricted (pressed) voice qualities. Note that the physician's normal "say 'ah' will not do. The oral cavity is open for ah, but the pharynx is too narrow. The vowel "ee" suffers from the high front position of the tongue, and "oo" has the disadvantage of a high back tongue position. The vowel (e) suffers from none of these limitations for photography through a rigid telescope, although "ee" gives the best results when one uses a fiberscope through the nose.

Plate 16, B, with breathy voice, has the vocal processes somewhat apart and there is a substantial chink at the posterior end of the folds, although the rest of the photograph is rather like Plate 16, A, with normal voice. This is not true for Plate 16, C, with tense voice, in which the false folds are almost together and the posterior cartilages are drawn medially and forward almost to the epiglottis. The relationship between the entry to the larynx and the surrounding lower pharynx is also changed. The xeroradiograms and the laryngeal photographs supplement each other very well.

Sound Spectrography

The interpretation of spectrograms is discussed on pp 1777 to 1784. Speech acoustics is a well-documented field, and the sound spectrograph enables the physician to examine the acoustic analysis of a segment of speech up to 2.5 seconds in duration, which is the length of a typical sentence. The printout provides "time" (0 to 2.5 seconds) on the horizontal axis and "formant frequency" (0 to 8000 Hz) on the vertical axis with "intensity" (that is, acoustic energy, or "loudness" in perceptual terms) on the third dimension of gray scale, that is, darker or lighter markings. If the sound spectrograph can be said to have a disadvantage, it is that more expensive spectrum analyzers are more precise and faster.

Figs. 96-21 to 96-23 are spectrograms of a speaker uttering "a bedege" with normal, breathy, and constricted voice qualities. Fig. 96-21, with normal voice, shows inter al, regular, and clear vertical striations, one for each vocal fold cycle, and regular, bold horizontal bands, each corresponding to a vocal tract resonance or "vowel formant". The noise produced by blowing across the top of an empty wine bottle is the resonance frequency of the air in that bottle. The third formants in Fig. 96-21 show signs of less-than-perfect voice quality. Fig. 96-22 with breathy voice shows markedly weak vertical striations even on the baseline (that is, zero Hz), where they are usually strongest. The absence of vertical striations indicates total lack of voicing. Light-colored striations indicate weak, breathy voice. The whole of this sentence should be well voiced. The formant bands are blurred and irregular because the resonance chambers are being excited by turbulent noise from the glottis rather than by glottal bursts.

Fig. 96-23, with constricted voice, shows irregularly spaced vertical striations because the folds are too tense to vibrate freely. If the duration of each glottal cycle is compared with the duration of the following cycle, and all of them with the mean duration, a variation of as much as 5% or 10% around the mean will correlate well with a subjective judgment of harshness. The vowel formants, too, are not bold and compact but ragged in appearance. Although the resonance cavities are being excited by voice, the irregular periodicity of that voice is not a good sound source.

Because "normal" speech typically contains some "non-normal" syllables and "disordered" speech some "normal" ones, the question of normalcy becomes statistical. As a result, the diagnostician will do well to record some standard text such as the 30-second "Rainbow Passage", and mark each syllable for any of about 30 acoustic features that seem deviant. The totalled percentage will give a probabilistic indication of a disorder that is valuable because each disorder is characterized by a distinctive set of features.

Glottography

Electroglottography (EGG) is carried out by strapping a collar around the patient's neck to position two small, flat electrodes on either side of the larynx with a third, ground electrode to the side of the neck. A current is passed through the larynx from electrode to electrode. The current passes easily when the vocal folds are together but with difficulty when the folds are apart because of the impedance to current flow caused by the change of medium from tissue to air. It might therefore be appropriately called an "impedance collar", although it is typically said to show "vocal fold contact area". To what degree it does in fact show contact area is currently a matter of some discussion, although it clearly does something of that kind. The printout takes the form of a rising and falling curve, with some workers preferring to put "open glottis" at the top whereas others put "closed glottis" there. The figures here do the latter. Some waveforms are almost sinusoidal. The most complex have six stages: (1) a horizontally directed completely open stage; (2) the slow-closing stage, which rises at less than 45 degrees; (3) the fast-closing stage, which rises very sharply, is always present, and typically accounts for only 10% of the duration of the whole cycle; (4) a near horizontally directed, completely closed stage that is only present in normal, strong, or hypertense voice; (5) the fast-opening stage, which is always present and falls steeply but less so than stage 3; and (6) the slow-opening stage, which falls at less than 45 degrees. There are at least 16 waveform types in theory, but in practive not all occur. In addition, waveforms may differ significantly in the percentages of the duration of the whole cycle occupied by a given stage.

Figs. 96-24 to 96-26 are glottograms made while a sustained (e) is uttered with normal, breathy, and constricted voice qualities. Each figure has a DEGG (vocal fold velocity) trace at the top and an EGG (vocal fold contact area) trace at the bottom. The DEGG positive peak indicates the moment when the vocal folds first come together; the EGG positive peaks show when both the upper and lower rims of the vocal folds are together; and the DEGG negative peak indicates the moment when the vocal folds finally move apart. Fig. 96-24, with normal voice, shows a stage 1 directly followed by a brief, fast-closing stage 3. The vocal folds stay together for some time and then open, first slowly and then fast. The closed time is almost 50% of the cycle. Fig. 96-25, with breathy voice, shows the vocal folds making rapid closing and opening gestures and then staying apart for two thirds of the cycle. Fig. 96-26, with constricted voice, shows a brief, open stage 1 directly followed by a brief fast-closing stage 3, as with normal voice. Then, however, the vocal folds not only stay together for some time but open rather slowly so that the closed time is two thirds of the cycle. All the figures were recorded at 2 msec/division and 500 mv/division.

These breathy and constricte utterances were nowhere near as breathy and constricted as is possible. They were at the fringes of "normal". The glottograph tends to perform porrly with very breathy voice and shows a very long closed stage and low amplitude for strained or strangled voice. At present relative amplitude on glottograms should probably be ignored.

On Subcategorization

We have so far examined and discussed data for three grossly different voice types in order to make the principal message clear: perceptual, acoustic, physical, and physiologic techniques are available for distinguishing between different modes of laryngeal function and each adds something to the interpretation of the others. However, it should not be imagined that physicians are restricted to interpreting only three gross types. We will conclude with a discussion of a wider range of glottogram waveforms.

It is a useful as a standard procedude to ask the patient to utter a sustained vowel (e) at high, medium, and low pitch, each one with high, mediuim, and low intensity. A 3x3, nine-cell, two-dimensional matrix is used for data collection. In addition, the trained professional voice user may utter each matrix with normal, breathy, and tense voice, so giving a 3x3x3, 27-cell three-dimensional matrix. Given a 27-cell matrix, plus some examples of falsetto and creaky voice at the two ends of the fundamental frequency spectrum, it is possible to collect a wide variety of EGG waveforms (Painter, 1988; 1990). The classification by stages described above may be subclassified by closed/open ratio, giving a total number of different waveforms equal to 16 basic types x 7 ratios = 112.

Eleven of these are easily distignuished and described even without quantification. (1) and (2) are almost sinusoidal or triangular-like waveforms associated with low-intensity utterances at any frequency. Falsetto voice also has rather sinusoidal waveforms; (3) is found with low-frequency utterances at medium and high intensity rather like Fig. 96-24, but the closed/open ratio is about 2:3; (4) is a medium-frequency, medium-intensity utterance with a 1:1 closed/open ratio; (5) are medium-frequency, high-intensity and medium-intensity, high-frequency utterances with a 3:2 ratio; whereas in (6) high-frequency, high-intensity utterances have a 2:1 ratio (Fig. 96-26); very tense voice (7) has a pronounced, fully closed stage 1 or 2; in (9) and (10) very low-frequency creaky and diplophonic utterances show unequal, double opening-closing gestures within each cycle. In one case, the second gesture is prominent; in the other case, it is rather small; in some cases, such as (11), they are equally spaced but differ in amplitude.

Summary

This section has described how four different instrumental techniques can throw light on three different voice qualities, but there are other techniques that have not been discussed (Hirano, 1981; Painter, 1979; Sundberg, 1987), and it has been shown (Painter, 1991a) that there are something on the order of 25 voice qualities. What the busy clinician needs is a protocol that will use these techniques to establish the parameters of vocal performance during the course of a patient's visit. Laryngoscopy is a well-established procedure for establishing a clinical diagnosis, yet it is a poor one for evaluating vocal performance. On the other hand, simultaneous physiologic and acoustic measures enable one to evaluate vocal performance but are not very useful for establishing a clinical diagnosis. This is because any category of voice problem will have a range of performance deficits, whereas several different problems may have similar performance deficits. Therefore the identification of lesions or poor speaking habits on the one hand and the quantification of vocal performance on the other are important considerations.

A protocol for voice evaluation should have a multi-channel input, should not be difficult to use, should take no more than 25 minutes to administer, and should provide the physician with a printout in standardized form before the patient leaves the room. The evaluation system used in the Voice Laboratory at Washington University in St. Louis meets these criteria (Painter, 1991b).

We began by demonstrating how a few different techniques can throw light on a limited number of voice qualities. We have concluded by extending this argument to the use of a multiparameter protocol that give quantified, comprehensive information on all voice qualities.

Electromyography of Laryngeal and Pharyngeal Muscles

Hajime Hirose

When a muscle is activated by neural impulses from the central nervous system, the electrical manifestation of a neuromuscular unit (NMU) potential accompanied by a contraction of the muscle fibers can be recorded as the electromyography (EMG) signal.

The merit of EMG is that it can provide valuable information not only about the pertinent muscle system but also about the function of the innervating nervous system, particularly the lower motor neuron.

In the part of this chapter by Harry Hollien the practical aspects of EMG of the laryngeal and pharyngeal muscles are discussed with special reference to clinical application.

EMG Recordings: Instrumentation and Techniques

When a neural impulse arrives at the end plate of a group of muscle fibers, a wave of depolarization sweeps along these fibers and a brief contraction is elicited. At the same time, an electrical potential dissipates into the surrounding tissues. These potentials can be picked up and displayed or recorded as EMG signals if a pair of electrodes are placed in the muscle tissue.

Electrodes

For recording EMG signals from the laryngeal and pharyngeal muscles, intramuscular electrodes must be used. These electrodes can be classified into two types: needle electrodes and wire electrodes. Needle electrodes are appropriate for clinical examination, in which the pattern of the firing or excitation of a single NMU at several different sites in a muscle must be examined. In most cases, bipolar concentric needle electrodes comparable in size to 23- to 27-gauge hypodermic needles are used.

Wire electrodes are useful for studying muscle activity with special reference to the kinesthetic function of the larynx or pharynx during speech or deglutition. The wire electrode system consists of a pair of wires threaded through a hypodermic needle. The tips of the wires are bent back to make tiny hooks. The hypodermic needle is used to guide the wire electrodes to the proper location, after which it is withdrawn to leave the wires hooked in the muscle tissue (Fig. 96-27).

EMG apparatus: electromyograph

For EMG data assessment, an electromyograph is used. It consists of an amplifier and a recording system usually equipped with a monitor oscilloscope and a loudspeaker. Many different types of electromyographs are commercially available at present, most of which allow the recording of EMG signals in the form of either a continuous trace or a single sweep.

Preparation for EMG examination

Sterilization of the electrodes is usually accomplished by antiseptic solutions. High-pressure heat can also be used in clinical practice.

If we adopt a percutaneous route for the electrode insertion, the skin is disinfected with an antiseptic swab. For peroral insertion, disinfection of the site of insertion is note made in general, but topical anesthesia is often administered to the pharyngeal mucosa in order to control the gag reflex.

Insertion techniques and verification of electrode placement

In principle, correct placement of the electrodes is verified by monitoring the muscle activity induced by appropriate gestures that have been considered pertinent for the contraction of the target muscle. For some muscles, however, there is still a lack of normative EMG data on which verification can rely. Also, in the case of paralyzed muscles, normal firing patterns cannot be obtained even if correct placement is made. Thus verification depends to a certain extent on the examiner's empiric judgment based on his or her knowledge of anatomy and clinical experience.

Intrinsic laryngeal muscles

A percutaneous route is taken for reaching the cricothyroid (CT), thyroarytenoid (TA), and lateral cricoarytenoid (LCA).

For the CT, the needle is inserted at the level of the lower edge of the cricoid cartilage and 5 mm lateral to the midline. The needle is then directed toward the inferior tuberculum of the thyroid cartilage. Verification of the correct placement is made by asking the subject to raise the pitch of his or her voice. The CT shows marked activity for a quick rise in voice pitch.

To reach the TA, insertion is made with the subject in a supine position and attempting sustained phonation. The skin is pierced at the midline at the level of the cricothyroid space. The needle is advanced to pass through the cricothyroid membrane and the tip is directed submucosally laterally and upward to reach the muscle. For verification, the subject is asked to sustain low-frequency phonation. The TA also shows marked activity during swallowing and glottal attack.

For the LCA, the point of insertion is almost the same as for the CT. The needle is then inserted laterally and slightly upward, penetrating the cricothyroid membrane at a point anterior to the inferior tuberculum of the thyroid cartilage and deeply enough to reach the LCA. Verification is made by having the patient hold his or her breath or attempt glottal stop production. These maneuvers, as well as swallowing, give rise to marked activity and serve to discriminate the LCA from the CT (Fig. 96-28).

The peroral route is taken to reach the posterior cricoarytenoid (PCA) and interarytenoid (IA). Under direct laryngoscopy, a curved probe bearing a needle electrode at its tip is directed to the point of insertion through the mucosa covering the targeted muscle.

For the PCA, the insertion is made into the belly of the muscle on the cricoid cartilage. Verification is made by having the subject repeat short periods of vowel phonation interspersed with deep, quick inspiration. The PCA is active for inspiration and suppressed during phonation.

For the IA, the insertion is made at the midline between the two arytenoid prominences under indirect laryngoscopy. Verification is made by asking the subject to produce a short period of phonation. The general pattern of IA activity is almost reciprocal to the PCA; there is marked activity for the period of phonation (Fig. 96-29).

Pharyngeal muscles

To reach the pharyngeal muscles, the peroral approach is always attempted with the subject in a sitting or a supine position. Wire electrodes are preferable, but needle electrodes can also be used for the levator palatini for clinical purposes.

For the levator palatini, insertion is made into the levator "dimple" on the soft palate, approximately 1 cm from the midline and 1 cm from the posterior edge of the hard palate, with the subject attempting sustained open vowel phonation. The tip of the needle is directed laterocranioposteriorly approximately 1 cm from the surface of the mucosa. Verification is made by asking the patient to repeat the (s) sound. Marked activity can be observed for this strong oral gesture if the electrodes are placed properly.

The palatoglossus and the palatopharyngeus are reached by inserting an angulated needle into the anterior and posterior pillars. Because the insertion is made under direct inspection, verification is satisfied if marked activity is shown when the patient attempts to swallow.

Normal EMG Patterns

In general, complete relaxation of healthy muscles gives no EMG activity, whereas NMU activity is recorded as action potentials when muscles contract. As the strength of muscle contraction increases, the number of excited NMUs also increases and forms an interference pattern.

The strength of the contraction of the abductor laryngeal muscle (PCA) increases during deep inspiration whereas that of the adductors increases during phonation, coughing, and swallowing. Pharyngeal muscle activity also increases during swallowing. The activity of the CT increases when the subject raises the pitch of the voice and is suppressed during deglutition. These is a clear reciprocal activity pattern between the abductor and adductor laryngeal muscle groups in voluntary actions in normal subjects.

Pathologic EMG Patterns: Laryngeal and Pharyngeal Paralyses

Clinically, EMG was developed in neurology as an improved diagnostic and prognostic tool. In otolaryngology, EMG is now widely accepted as a routine procedure for the examination of facial, pharyngeal, and laryngeal paralysis. In particular, EMG is considered very useful in differentiating laryngeal paralysis from mechanical fixation of the cricoarytenoid joint and in estimating the degree and prognosis of the paralysis.

In clinical situations, a percutaneous EMG examination using a bipolar concentric needle electrode is preferable, and in the case of laryngeal examination, the two innervation territories (that is, the TA innervated by the recurrent laryngeal nerve and the CT innervated by the superior laryngeal nerve) are examined bilaterally.

Fig. 96-30 shows an example of normal laryngeal EMG patterns of the TA during phonation and deglutition. In both maneuvers, normal voluntary activity obtained from the TA is characterized by an interference pattern. The activity is more marked during deglutition than in phonation, because the former requires a stronger laryngeal constriction.

Fig. 96-31, A, shows an example of the EMG pattern in incomplete laryngeal paralysis. In this case, volitional activity is partially preserved. Even if laryngoscopy reveals immobile vocal cords in such cases, the recovery of vocal cord mobility can be expected, provided that the existence of some remaining volitional EMG activity can be confirmed in the early stages of the development of the apparent analysis.

Fig. 96-31, B, shows an example of a sign of complete denervation, in which no volitional activity is recorded but some small involuntary potentials are seen. This type of discharge is termed a fibrillation potential. If such findings indicating complete denervation are obtained, the prognosis for the paralysis is usually very poor.

It is often the case that volitional EMG activity is recorded months to years after the onset of paralysis without any sign of a recovery of vocal cord mobility. In such cases the apparent immobility of the vocal cords is considered to be caused by a so-called misdirected regeneration, in which a confusion of the regenerating nerve fibers occurs in the differentiated innervation between the abductor and adductor muscle groups. It may also be possible that an increase in the amount of intramuscular connective tissue associated with the atrophy of muscle fibers results in an abnormal contraction of the reinnervated muscle.

Limitations of EMG Examination

The limitations of EMG assessment of pharyngeal and laryngeal function partly lie in the technical difficulty of its application. Particularly ni the case of paralysis, verification of the accurate placement of electrodes is notably difficult.

In the case of nerve paralysis, the interpretation of recorded data is often difficult in terms of the prognostic evaluation, particularly in cases of long-standing paralysis.

From the physiologic viewpoint, it is often argued that the sampling size of the active motor units is too small to represent a given muscle as a whole. Also, the relationship between measured EMG activity and its mechanical effect is not necessarily linear; therefore quantitative descriptions of the obtained results are often difficult to interpret.

If these limitations are taken into consideration, the use of EMG can provide useful information for both clinical practice and research in otolaryngology.

Inferring Laryngeal Characteristics From Phonatory Output

Harry Hollien

The physician is accustomed to assessing nonnormal morphology in the larynx but typically has much less experience in judging deviant laryngeal function. Further, although the sex and age of the patient are easy to ascertain, normative data on the speech characteristics of the sexes in different age groups and in varying states of health are still not fully documented. However, lately substantial strides have been made in the development of methods that permit practitioners to assess characteristics such as the gender, age, and health of persons by measurements of human phonation. The techniques and procedures that permit this assessment have been developed by investigators from a number of disciplines; however, the contributions made by phoneticians and engineers are particularly extensive. Of course, the source of a specific technique is inconsequential; what *is* important is that it (1) is valid for the purpose; (2) behaves reliably, and (3) is standardized on the populations for which it is intended. In short, each technique - whether machine or human, whether computer supported or descriptive - must be calibrated. It is not sufficient for a procedure to be proposed, or even used; it must be thoroughly evaluated before application. Hence, this part of the chapter includes both the strengths and weaknesses of the approaches discussed.

What are the voice protocols that permit evaluation of the health, age, and sex of a speaker, and what stable information is available in these areas? Generally, two approaches are available and have been successfully applied. One utilizes the assessment of the perceptual responses made by human listeners to vocal stimuli; the other involves a variety of machine processes that identify or quantify specific vocal parameters. To thoroughly review all of the techniques available and evaluate and codify all of the data relevant to the cited issues is beyond the scope of this chapter. However, a briew overview of these elements is presented below.

Vocal Percepts

Evaluating available data

Using humans as measurement devices is an ancient and honorable research technique. However, it can be easily abused. Basically it requires that specific stimuli are chosen and presented to auditors in a manner that will permit useful responses. The question asked, the development of the stimuli, and the experimental procedure employed are all important. All too often an investigator asks intelligent questions but does not design research to answer them; or the stimuli employed are indirectly misleading. To illustrate, spurious sounds associated with specific stimuli can provide biasing information to the listener. Fatigue or boredom also can bias results; even the instructions given can do so. Stimulus presentation is also important. For example, paired comparison techniques are quite powerful unless there are so many stimuli that subjects begin to react negatively. Totally "blind" procedures are excellent (for example, "you will hear twenty sounds; please sort them into two groups by placing the number of each in column A or in column B"). Thus it is important to carefully examine the hypotheses and procedures used in any particular study for flaws that might invalidate the data. Such errors can be and are made. To illustrate, some years ago a seemingly impeccable scientist published an article on dolphin hearing in a prestigious journal. He indicated that the hearing of these aquatic mammals had a range up to an exceeding 150 kHz; he did so even though the underwater sound source he employed in the research had an upper frequency limit of 20 kHz. In any case, no investigator's work should be exempt from scrutiny.

It is also important to evaluate the appropriateness of the listeners used in any perceptual experiment. Are they appropriate for the task? For example, before the experiment the investigators should have assessed the auditors' (1) hearing, (2) background, (3) ability to respond properly, and (4) place on the "learning curve". Finally, consideration of the mechanical aspects of test administration, data reduction, and testing (including statistical) should be included in your assessment of the validity of the data you are presented. It is only at this juncture that you as a physician can begin to discover if the materials you have reviewed are valid, realiable, and useful to you for your particular purpose. In short, the physician, that is, the "consumer", should evaluate the effectiveness of the investigators and their work.

Using perceptual data

The use of data based on the perceptual resopnses to phonation depends on several relationships. Fist is the validity of the entire process. More important is the issue of whether or not the information provides answers to the clinical or educational questions being asked. Moreover, although not always obvious, if a practitioner is to assess or understand the nature and consequence of a pathologic condition, it is first necessary to know what is normal for his or her patients - or at least, what is not pathologic. Thus perceptual data about health should always include the characteristics of non-pathologic populations or, if the focus is on a particular disease, the data for normal persons should be referenced as a baseline. The two examples below illustrate the positions cited above and provide summary information about certain phonational characteristics. A third example is included to illustrate how perceptual data may be validated by machine procedures. In any case, the three examples cited include (1) the estimation of age from voice samples, (2) how an nonnormal phonatory percept (hoarseness) can be standardized as a test for clinical use, and (3) the combination of perceptual and acoustic evaluations to permit the accurate estimation of phonational frequency ranges (PFR).

Perception of age

Some years ago Shipp and Hollien (1969) reported a study in which 175 male talkers - 25 per decade, 20 to 89 years of age - were carefully selected and then required to read a standard passage. Two sentences (22 words each) were drawn from the 175 samples, randomized and played to 3 (adequate-to-task) groups of 25 to 40 young adult listeners. Three scaling techniques were used in judging the ages of the talkers: a 3-point scale, a 7-point scale (decades 2 to 8), and a direct (age) magnitude procedure. All three produced essentially the same results. Mean age estimates progressed in an orderly basis from 28 years for the 20- to 29-year-old group to 66 years for the 80- to 89-year-old group. To quote Shipp and Hollien, these results permitted quantification of the "empirical impression that most people are able to estimate a talker's age from his voice". Although these data have been verified, two cautions must be noted. First, although the investigators report a chronologic/perceived age correlation of $r = + 0.88$, plots revealed soem scattering among the data. For example, at least one 20-year-old was judged to be 44 years of age and one 80-year-old as only 60 years of age. The spread among listener responses was similar. Hence, it must be said that s single auditor judging the age of a single talker might exhibit considerable error. Second, the auditors in that study were young adults and their estimation errors for the older talkers were greater than for younger. Because good data are not yet available about the effects of age bias on age judgments, the reason for this truncation of the age estimates for the older groups is not known. Nevertheless, it must be concluded that, although practitioners may be able to accurately judge the ages of a large number of talkers, they also may experience a substantial problem with a specific person - and that their age probably will bias their judgment.

Clinical precepts

It is now apparent that the data gathered from most perceptual studies suggest that this process is a very complex one and that, although certain classes of professionals can accurately classify *groups* of talkers on a variety of continua, it is difficult to accurately classify a single talker. Consider the study currently being completed by Anders et al (in press). Here the focus is on the following queries: can the presence or absence of hoarseness be accurately determined and can its severity be validly scaled? To answer these questions ni a somewhat more sophisticated manner than is usual, 11 groups of 8 to 12 listeners (different types of professionals and controls) from three countries were asked to evaluate a single set of 40 phonatory samples prestructured into the four categories of (1) normal and (2) mildly, (3) moderately, and (4) severely hoarse voice. It was found that all groups, including the controls, could make the normal versus hoarse distinction at levels varying from 74% to 93% accuracy; that hoarseness severity could be correctly estimated at levels significantly above chance, and that there was a cross-cultural effect, with the American groups the most permissive and the Finnish groups the most severe in their judgments.

Of what importance are data such as these to practitioners? First, they illustrate that certain vocal qualities are so robust in nature that universal (or nearly universal) definitions can be developed - and, given a standard definition, most people who hear that particular quality will identify it in the same way. It also demonstrates that such variables as orientation, training, culture, and language can operate to modify a concept or operation of this type. Perhaps most important, it permits practitioners to develop standards for diagnostic and treatment protocols that in turn allow them to communicate effectively with other relevant professionals.

Phonational frequency ranges

Up until the 1950s, it was difficult for practitioners and scientists alike to think about or measure the "range" of voice. Some workers attempted to use "singers' ranges"; other applied ranges based on vocal registers. Unfortunately, however, not much data were available on singers' ranges and in any case the voices of singers do not necessarily parallel either those of normal persons or those exhibited by persons with voice pathologic conditions. Moreover, at that point in time (the 1950s), even the concept of voice registers was open to serious controversy. In response, phoneticians began to investigate voice registers and were able to structure them in a reasonably organized manner (if in a somewhat different way than did singers). Based on the vocal register model, a method of defining voice ranges was developed at the University of Iowa, stimulated by Grant Fairbanks and James F. Curtis. Specifically, it was postulated that there are two to five voice registers and that two of them are particularly relevant to speech (for in-depth discussions of vocal registers, see Hollien, 1974 and 1983). The use of two registers, rather than a single one, to determine a person's voice frequency range resulted from the fact that the lower boundary of the modal register could be easily and accurately determined, but its upper limit could not - and that the converse was true of the next higher register (falsetto). Thus even though most of the fundamental frequencies ordinarily used in speaking and singing could be found in the lower half of a person's modal register, the total extent of the modal an falsetto registers had to be used as an index of the physiologic (voice) range (as these registers could be validly measured only as a pair). Hence the cited dimensions began to be used in voice research, speech therapy, and related areas. But several questions remained: How can the practitioner use concepts such as these? What can be expected of a given person? What is normal for children, adults, or the aged?

In response to these questions, Hollien et al published a paper in 1971 in which they defined phonational frequency range (PFR) as the extent of the modal and falsetto registers combined; further, they provided baseline data for normal adult men and women. Specifically, these investigators had 332 men and 202 women (aged 18 to 36 years) phonate frequencies corresponding to those of the equal tempered music scale. It was found that both men and women exhibited a slightly greater than three octave PFR (38 and 38 ST for men and women, respectively) with the ranges of nearly all subjects falling between two and four octaves. But how stable/valid are these data? Hollien et al retested 6% of their subjects and found a mean error of just 1.3%. Still later the procedure was validated by a machine procedure on 11% of the subjects plus 32 additional persons (20 men and 12 women). The machine procedures involved the wave-to-wave measurements of oscillograms (actually phonellograms) in order to determine the mean frequency *actually* phonated at the PFR extremes (that is, the two registers). The difference between the perceptual and machine procedures was 0.57 ST (the machine values were greater, of course). Because addition of this value to the raw data did not result in changes in any of the means, the difference was judged trivial. More importantly, here was a case in which a machine approach validated a perceptual one. In any case, Fig. 96-32 provides data on the dispersion of PFR among adult men and women. The practitioner can utilize data such as these to establish limits to normal phonation in these populations. Deviations - especially those on the "low" side - can be considered as indicative of a laryngeal pathologic condition. Second, data are available that demonstrate that adolescents' phonational frequency ranges measured during pubescent voice change (AVC) are generally greater than those for adults and that there is little or no evidence to support the notion that markedly restricted PFR correlates with AVC (Hollien, 1978). Thus if a particular adolescent exhibits a sharply reduced PFR, the cause of the conition probably is a laryngeal pathologic condition. Finally, PFR data are published on children and the elderly, practitioners will be able to establish normal-nonnormal PFR standards for general use in diagnostic evaluations and will also have guidelines to assist them in the evaluation of the success of their therapy.

Instrumental Evaluation of Phonation

Not long ago nearly all of the apparatus used in the study of speech and voice was mechanical. The first upward shift came during the middle of this century. At that time, good electronic equipment began to be developed and used in place of mechanical devices. As would be expected, this electroacoustical equipment was not just more complex; much of it also was quite versatile and certainly more suitable for voice research. Just recently, a second revolution has started to take place, involving a shift from hardware to software (computers). This revolution is far from over. Presently three classes of instruments are available to the investigator: (1) harware - that is, electronic devices that can perform a large variety of functions and do so independently of other equipment; (2) hybrid systems, those in which electronic gear is coupled to a computer and they work symbiotically to perform some sort of combined operation; and (3) computers, which are programmed to simulate some electroacoustical (or related) function so that appropriate data can be developed and evaluated. All three of these approaches are considered below.

The systems that can be applied to the study of voice are so many and varied that it is difficult to organize them into meaningful categories. One way by which appropriate structuring might be accomplished is to fit all of the machines and approaches into the four domains that are used to describe the speech/voice signal: (1) fundamental frequency (perceived pitch); (2) wave composition or spectra (quality); (3) intensity (loudness); and (4) time or temporal features. However, because an entire book could be written about the devices in question - or the four psychophysical categories - only a very small sample from among them will be selected for discussion.

Speaking fundamental frequency

Almost every *periodic* sound, including human phonation, is accompanied by a "pitch" that can be perceived by the auditor. As it turns out, the physical correlate of pitch, fundamental frequency (F0) or speaking fundamental frequency (SFF), is an acoustic parameter that would appear easy to extract and measure but in reality is not. Because F0 in voice/speech is a direct product of the vibratory rate of the vocal folds, it would seem that any device that could capture this pattern, then store and measure it, would meet the requirements for obtaining F0. Direct observation of the vocal folds is possible of course but, unfortunately, it is impractical as an F0 measurement technique. Oscillographic techniques, while easy to understand and use, also exhibit problems of a rather substantial nature. First, speech/voice waves ordinarily are so complex that it is quite difficult to determine where a particular wave starts and where it stops. Second, calibration of these systems is tricky; thus establishing a stable time base (so that the measurements can be accurately converted to frequency) is difficult. Worst of all, because only hand or interactive measurements are possible, data reduction is extremely time consuming.

Sometimes investigators use equipment not designed to provide F0 measurements in an attempt to do so; the use of sound spectrographs of the time-frequency-amplitude type is a case in point. Here the SFF "data" are obtained by estimation processes; that is, they are generated by a scheme such as periodic measurements of the tenth harmonic partial, which often can be seen when a spectrograph is operated in its narrow band mode. Presumably means SFF can then be obtained by dividing the average frequency of the tenth partial by 10. There are other approaches, of course. Most are time consuming and not particularly accurate. In any case, it can be seen that it is rather difficult after all to obtain SFF; the reasons that underlie the problem may be found in Hollien (1981).

Automatic, semiautomatic, or computer analysis of voice period is desirable if this vocal characteristic is to be studied efficiently; several are now available. Some are designed for the production of general SFF information, others for specialized purposes. They inclue the "Purdue pitch meter" (Dempsey et al, 1950), a several-stage device that has been utilized for some years (Mysak, 1959), and a system developed by Böe and Rakotofiringa (1971) that has been used to support a number of their research projects. In both instances, the device utilized was developed primarily (but not exclusively) to permit analysis of reasonably large groups of subjects for the purpose of describing SFF usage in normally speaking persons. Although the validitiy of these systems has been established, data reduction can be somewhat tedious.

Another approach - best typified by the equipment of Frøkjaer-Jensen or by the Kay Elemetrics Visi-Pitch 6087 - attempts to provide a visual display of F0 produced by persons who exhibit either normal or abnormal voices. Even though both of these units have digital capability, they are best used in teaching or in clinical evaluations of voice disorders. Analysis of large groups of subjects is cumbersome when these devices are utilized in SFF research.

Finally, there are computer approaches for F0 extraction. Although the software here is rather sophisticated, these approaches usually require extensive processing time and produce rather small amount of data. Moreover, as with most devices of this class, they *sample* F0 rather than measure every wave in a speech segment. Indeed, as McKinney pointed out back in 1965, there are very few systems available that operate with both satisfactory validity and high efficiency.

One system that meets the cited criteria is available. Indeed, much of the published quantitative data about SFF has been processed by this particular machine (or its antecedents). It is the IASCP Fundamental Frequency Indicator (FFI-8), a hybrid device that, when coupled to a PDP-11/23 computer, measures the F0 of speech/phonation (Hollien, 1981). Briefly, FFI-8 is designed to automatically extract F0 from complex waves. To do so, it does not sample the speech signal, rather, it measure every periodic wave in any given series. Basically, the system consists of a set of one-half octave, high-resolution, low-pass filters coupled to high-speed switching circuits. The filters serve to separate the fundamental from its harmonic overtones. The high-speech relays operate to shut off all filters above the lowest one containing energy; hysteresis is built in to circumvent chatter. The extracted fundamental is squared and fed to the computer, which collects, stores, and processes these materials. Addditional software permits serial measurements to be made, appropriate metrics calculated, and statistical procedures applied. Typical of the obtained metrics are (1) geometric mean frequency level, (2) the standard deviation of the distribution, and (3) histograms of the distribution (other metrics are also available). A schematic FFI-8 may be seen in Fig. 96-33; its printout is shown in Fig. 96-34.

A device such as FFI-8 can be used to answer questions about the SFF of infants, children, adolescents, adults, middle-aged adults, and older people. For example, the SFF correlates of puberty, menopause, and/or the aging process can be calculated. In 1976, Hollien and Hollien were able to point out that the gender confusions that sometimes occur when preadolescent children are heard but not seen may result from the fact that girls apparently have a *lower* SFF than boys. As it turns out, there is a relatively simple reason for this relationship. During the developmental years, girls, are physiologically somewhat more advanced than boys; hense it is the mechanical substructure of these populations that permits the cited observation to be made.

As has been implied, FFI data can be utilized in the study of the voice change process that accompanies puberty. For example, in 1979, Doherty and Hollien reported a new analysis of the 14 voice, age, and physical size parameters collected over a 6-year period on 48 adolescent boys. Using a massive compilation of data, a revised model predicting preadolescence, neoadolescence, and postadolescence was developed. To be specific, Doherty and Hollien were able to use a cluster analysis statistical technique to test Hollien's previous model, which had indicated that on the average the pubescent period in men lasted from 12.5 years (preadolescence) to 15.0 years (postadolescence) and that the SFF limits for the preadolescent, neoadolescent, and postadolescent voice should be 190 Hz and above, 150 to 189 Hz, and 149 Hz and below, respectively. The computed model provided a pubescent range of 13.0 to 14.5 years and SFF cluster means of 246, 217, and 142 Hz.

Finally, SFF curves for adults can be generated; they are based on the work of Benjamin (1981), de Pinto and Hollien (1982), Gilbert and Weismer (1974), Hollien and Paul (1969), Hollien et al (1982), Linke (1973), McGlone and Hollien (1963), and Stoicheff (1981) for women and Benjamin (1981), Hollien and Jackson (1973), Hollien and Schipp (1972), and Mysak (1959) for men. These curves are presented in Fig. 96-35. Please note that they also provide an estimated SFF *range* for each sex and age combination.

Vocal jitter

Closely related to speaking fundamental frequency is vocal jitter, a laryngeal phenomenon defined as the cycle-to-cycle variation found within successive periods of laryngeal vibratory pattern. Because jitter is well accepted as a common occurrence in normal phonation, virtually all vocal signals can be considered to be "quasiperiodic" in nature (Flanagan, 1958; Fant, 1960; Peterson and Shoup, 1966). Moreover, the laryngeal phenomenon of vocal jitter has been found to exist under a variety of phonatory conditions: (1) in sustained phonation (Beckett, 1969; Hollien et al, 1973; Horii, 1979), (2) in connected speech (Lieberman, 1961, 1963), and (3) as related to unusual laryngeal conditions and voice pathologic conditions.

Traditionally, two basic measurement techniques have been applied to the study of vocal jitter. In one case, data have been extracted either from ultra high-speed motion pictures of the waveform (Lieberman, 1963; Moore and Thompson, 1965), photographs of the (ultra high-speed) acoustic signal displayed on an oscilloscope (Hecker and Kruel, 1971; Hollien et al, 1973), or a similar system (Beckett, 1969). More recently, computer-based jitter extraction systems have been implemented by Hollien et al and by others (Anderson et al, 1976; Davis, 1976; Doherty, 1977; Horii, 1979). In these cases algorithms underlying the computational process are applied to the digitized representation of either the acoustic waveform or its residues. The procedures utilized are based on some metho of "peak picking" or on autocorrelation techniques. For example, Davis (1976) has described a method that seeks the highest correlation between the two sets of points that define first one period and then the next. In addition, Doherty (1977) recently has proposed a system based on modified use of the Fundamental Frequency Indicator (FFI-8). As a follow-up to Doherty, Hollien et al have developed software that permits continuous computer measurement of any series of waves at sampling rates of 100K; statistical comparisons are made between all adjacent waves, and a jitter metric is developed. This technique has the advantage of being faster than the other computer approaches cited and just as accurate.

But what useful information can vocal jitter research provide the practitioner? For one thing, increases in the magnitude of jitter suggests the presence of laryngeal tumors such as those associated with nodules or cancers (Hartman et al, 1982; Lieberman, 1961; Ludlow et al, 1983). At the very least, abnormal jitter signals the presence of a voice disorder. Moreover, elevated vocal jitter or, at least, roughness or tremors, appear to be associated with old age, although the relationships here are not conclusive. The basic problem with jitter-related metrics is that so little research has been carried out on this entity that most potential relationships are speculative.

Spectral analysis approaches

Spectral analyses are used for many purposes. For example, short-term displays correlate with the vowel being produced, and *long-term* spectral configurations are thought to relate to general voice quality. Further, the structure and nature of a sound spectrometer determine the amount and quality of the information it makes available. As would be expected, there are many types of sound spectrometers; basically, three types or classes dominate the others. They are (1) "instantaneous" spectrometers, (2) long-term units, and (3) time-frequency-amplitude (t-f-a) spectrometers. Although the use of t-f-a spectrograms is fairly widespread, this technique is considered to be of marginal scientific significance. Hence, only the first two types of spectrographs are reviewed below.

"Instantaneous" spectrometers and wave analyzers

A device of this type repeatedly and continuously sweeps the input signal to record and portray the energy found at each harmonic partial throughout the wave. Because it takes time to make a spectrogram of this type, and ordinarily, a person cannot sustain phonation with an exact F0 and spectrum, the product of a wave analyzer is not precisely instantaneous in nature; often the output will be a little "blurred". Nevertheless, reasonably accurate calculations for steady state tones (or at least, quasiperiodic phonation) can be achieved by this method. Moreover, commercially available units usually provide interface electronics so that the output of the wave analyzer can be fed directly to a computer and processed digitally. In any case these devices are used to study the spectral components associated with various voice registers, with singing, with voice disorders, and so on. Although some interesting information has emerged from such research (for example, the number of partials in modal and falsetto phonation are similar, but the patterns are different), few if any definitive relationships have been proposed or uncovered. Other types of equipment within this class include Fast Fourier Transform spectrometers, which are quite versatile as their operational modes range from wave analysis to power spectra.

Long-term spectral analysis

The overall spectrum of an utterance can be represented by the summed-over-time frequency/intensity relationships of the individual partials within an acoustic wave. That is, a power spectrograph measures all the energy contained in a signal at each of its frequencies over a relatively long period of time; it then displays this information either digitally or graphically as a frequency-amplitude envelope. The curve produced provides either a profile for an individual patient or a composite profile for a group of talkers.

Power spectra (and related measures) have been used for many purposes. For example, the technique has been utilized to determine the identity of specific persons from samples of their speech, and there have been suggestions in the literature that this type of analysis could be useful in predicting a person's age or emotional state (that is, if the person is suffering from some type of psychosis). Unfortunately, research that would provide definitive data about these issues has yet to be carried out.

On the other hand, consider the data presented by Wendler et al (1980). These investigations report on a study in which they took 40 samples of normal and hoarse speech (protocols for this research are reviewed above; see Anders et al) and subjected them to long-term spectral analysis. They used the long-term spectral analysis (LTS) technique in an attempt to validate the diagnostic judgments of practitioners and to differentiate among various levels of vocal roughness and breathiness. Discriminative analysis of statistical techniques were applied to the resulting data sets using posterior classifications. These investigators report that all the normal subjects and subjects with mild hoarseness were correctly classified by the LTS technique. The correct identification of hoarseness and the scaling of severity for the moderately and severely hoarse groups ranged from 91% to 92%. Further, the roughness and breathiness classifications approximated those for scaled hoarseness. Thus it can be seen that relatively sophisticated instrumental techniques now can be used in the support of research on laryngeal health and age. LTS is among them. It is suggested that continued use of this technique could lead to quantitative data on a variety of voice pathologic conditions.

Computers

Although many other types of instruments could be described, most are yet to be adapted for easy application to research of an applied nature. Not so for computers, as the many references throughout this chapter will demonstrate. In some cases the measurement techniques discussed were composed totally of software, but in other cases the system was a hybrid; in nearly all instances the software was adapted to the project. Indeed, it must be said that the modern digital computer is advancing rapidly in both sophistication and versatility. It is a tool that will permit the practitioner and scientist to think quantitatively about pathologic conditions and study them in an organized manner. However, two facts about computers must be kept in mind.

First, it should be remembered that the digital computer is simply a tool. Accordingly, research or assessment should not be adjusted to fit the computer or the available programs; rather, computer capability, and associated software, should be developed to serve the needs of the tests or procedures you are carrying out. Moreover, it should be remembered that computers as yet have but a limited usefulness for many types of speech and voice research. That is, although it is possible for investigators in a discipline such as psychoacoustics to design research in which the computer controls the experiment and then processes the subject's responses, an approach of this type is not yet possible for most speech research. Indeed, much of the current work, where attempts are made to relate various speech/voice parameters to behavioral states, consists primarily of exploratory ventures.

The above cautions should not be considered to suggest that computers are not powerful tools in the instrumental repertoire of the scientist or practitioner interested in the study of phonatory behaviors. The modern digital computer permits the efficient and effective storage and processing of data; it also allows the investigator to be privy to the results of his or her work within a short period of time. Finally, computers are now permitting the mathematical study of speech behaviors not previously possible and the construction of theoretic models of phonation (see Hollien et al, 1982). Fig. 96-36 illustrates just such a model. Here data from many instrumental and perceptual experiments have been combined and analyzed to permit generation of predictive curves for four speech/voice behavioral changes that occur as a function of induced stress (at least, for the average person). In summation, it can be said that the impact of computers on modern technology and science is substantial. Their intelligent and judicious use should greatly enhance research in the field being reviewed.

Spectrographic Evaluation of Voice Characteristics

Satoshi Imaizumi

Spectrographic Parameters for Evaluation Laryngeal Function

The sound spectrograph is a useful tool with which to analyze laryngeal function in speech. The acoustic parameters observable from sound spectrograms and relevant for the assessment of vocal function can demonstrate one or more of the following aspects of the vocal signal:

1. Fundamental frequency (F0)

a. The mean F0 of a given phonation

b. The possible range of F0 of a given subject

c. Fluctuation in F0 for a given phonation

2. Amplitude or intensity of the acoustic waveform (I)

a. The mean I of a given phonation

b. The possible range of I for a given subject

c. Fluctuation in I for a given phonation

3. The amount of energy in the high-frequency components

4. The amount of energy in the noise components.

Display Formats of Sound Spectrograph

The sound spectrograph analyzes an input voice signal and presents it in a variety of display formats. Among these, the following three formats are most fundamental and useful for the assessment of laryngeal function.

Three-dimensional display of power spectrum

The spectrograph resolves an input voice signal into its spectral components and displays them on a three-dimensional pattern, that is, time versus frequency versus level pattern. Examples are shown in Fig. 96-37. The analyzed voice sample is sustained vowel /e/ uttered by a male speaker with normal larynx. The x axis represents time in seconds (sec), the y axis represents frequency in kHz, and the z axis represents the level of power spectrum in decibels by varying the darkness of the display. A darker portion in the display has a stronger level of power spectrum.

The display in Fig. 96-37, A, is made using an analysis filter with the bandwidth of 45 Hz (narrow-band filter). In this display, the fundamental frequency and its harmonics can be observed clearly. On the other hand, the display in Fig. 96-37, B, is made using an analysis filter with the bandwidth of 300 Hz (wide-band filter). This display shows regularly repeated glottal pulses.

The use of an analysis filter with narrower bandwidth brings higher frequency resolution. On the contrary, the use of an analysis filter with wider bandwidth gives higher time resolution.

The patterns shown in Fig. 96-38 show the analysis results for a pathologic voice sample, /e/. In Fig. 96-38, A, the fundamental frequency fluctuates irregularly and the harmonics are not clear when compared to Fig. 96-37, A. In Fig. 96-38, B, glottal pulses are irregular.

As exemplified in Fig. 96-38, the pattern is useful to observe the fundamental frequency, or pitch periods, of vocal vibration and also the harmonic structure of voice waveform. Pathologic voices tend to reveal irregular pitch periods and unclear harmonic structures. The pattern can also be used to measure the possible range of F0 and its approximate mean value.

Amplitude display

The amplitude display produces a single trace representing the amplitude envelope of the analyzed signal as shown in Fig. 96-37, C, and also in Fig. 96-38, C.

The trace of the amplitude display shown in Fig. 96-37, C, has a steady portion, whereas it fluctuates irregularly in Fig. 96-38, C. Pathologic voices tend to have fluctuating amplitue envelopes. The amplitude display can be used to measure the possible range of the intensity and its approximate mean value.

Sectioned display of power spectrum

The section is a two-dimensional display of power spectrum at a particular time point of interest. Examples shown in Figs. 96-37, D, and 96-38, D, are produced using a narrow-band analysis filter, and those shown in Fig. 96-37, E, and Fig. 96-38, E, are produced using a wide-band analysis filter.

The section of Fig. 96-37, D, shows clear harmonic peaks, whereas that of Fig. 96-38, D, reveals nonharmonic components (noise) that are as loud as the harmonics. Pathologic voices generally tend to contain louder noise components than normal ones.

The amount of high-frequency component can be observed also from section displays. Some pathologic voices, especially those perceived as "strained", may contain a larger amount of high-frequency component. On the contrary, some "asthenic" voices reveal a small amount of high-frequency component.

Section displays produced with a wide-band filter can be used to observe formant frequencies, overall spectral shapes, or both.

Methods for Quantitative Measurements

The methods for quantitative assessment of laryngeal function mentioned below were introduced by Imaizumi et al (1980). They originally measured eight acoustic parameters from sound spectrograms of sustained vowel /e/. The frequency range between 80 Hz and 5 kHz was analyzed. Among these parameters, the following four items are relatively easy to measure and useful for differentiating pathologic voice samples from normal ones.

Extent of fundamental frequency fluctuation

The extent of fundamental frequency fluctuation is measured on a narrow-band filtered pattern. The extent of fluctuation is defined as the percent score of the ratio of the peak-to-peak value of fluctuation (deltaF0) to the mean fundamental frequency (F0). The actual measurements are made with a harmonic component as high as possible for the sake of accuracy as shown in Fig. 96-39, A. In this figure the extent of fluctuation is calculated at 100 x (width of fluctuation in Hz/approximate mean frequency of the harmonic component in Hz).

The extent of fundamental frequency fluctuation tends to be larger for pathologic voices, especially for ones perceived as "rough", than for normal ones.

Extent of amplitude fluctuation

The extent of amplitude fluctuation is defined as the peak-to-peak value in dB measured on an amplitude display as shown in Fig. 96-39, B. If the amplitude display contains fine ripples corresponding to the pitch periods, fluctuation in the peaks of the ripples should be measured. This occurs according to the sound spectrograph used.

The extent of amplitude fluctuation tends to be large for pathologic voices, especially for one perceived as "rough", when compared with normal ones.

Relative level of high-frequency components

A section display is used for the measurement of the relative level of the high-frequency components, which is defined as the ratio of the average intensity level of harmonics within 3.5 and 4.5 kHz to that below 1 kHz. This can b calculated as the difference of two mean values: the value averaged over the harmonic peaks within 3.5 and 4.5 kHz and that averaged over the harmonic peaks below 1 kHz. Two examples are shown in Fig. 96-39, C. In this figure, the spectral envelope A(F) obtained by connecting the peaks of the harmonics is used to calculate the relative level of high-frequency harmonics. If the peaks of the harmonics are clearly recognizable, those peaks can be directly used to calculate this parameter.

The voices perceived as "asthenic" sometimes reveal a very poor relative level of higher frequency component. On the contrary, some "strained" voices show a very high relative level of the high-frequency components.

Relative level of noise

Relative level of noise is defined as the ratio of the noise level to the level of the harmonics and is estimated on a section display. This parameter is calculated as the difference between the two mean values within a range between 2 and 3 kHz: the value averaged over the peaks and that averaged over throughs. Two examples are shown in Fig. 96-39, D. In this figure, two envelopes, A(F) and B(F), are used to calculate this parameter. A(F) is the spectral envelope obtained by connecting the peaks of the harmonics, and B(F) is the envelope obtained by connecting the throughs. If the peaks of the harmonics and the throughs between those peaks are clearly recognizable, this parameter can be simply calculated as a mean value of the differences between the peaks and the troughs within the specified frequency range.

Pathologic voices, especially ones perceived as "breathy", tend to contain a larger amount of noise.

Application for Medical Use

Imaizumi et al (1980) investigated the possibility of differentiating causative diseases using sound spectrographic analysis of pathologic voice samples. They analyzed sustained vowel /e/ recorded from 130 subjects, consisting of 20 normal subjects and 110 cases of laryngeal cancer, recurrent laryngeal nerve paralysis, vocal polyp, nodule, or sulcus vocalis. The spectrographic parameters mentioned above and several others were measured. After the multi-dimensional statistic analysis of those parameters, they reported that the voice samples of about 85% of the cases were differentiated from those recorded from the normal subjects. They concluded that the spectrographic analysis is useful for evaluating the change in voice qualities caused by laryngeal pathologic conditions.

Choi et al (1980) also conducted spectrographic analysis of pathologic voices. They found that the first formant frequency is also useful as one of the parameters for evaluating laryngeal function.

Kim et al (1982) reported spectrographic characteristics of the voices recorded from patients with recurrent laryngeal nerve paralysis, and Yoon et al (1984) reported their results on patients with glottic carcinomas. Fujiu et al (1988) presented an improved method for the measurement of the relative noise level by means of a digital sound spectrograph.

For the spectrographic assessment of laryngeal function, qualitative observation based on a visual inspection can also be useful, as shown by Yanagihara (1967). Some other methods of acoustic analysis for assessing phonatory function are reviewed by Hirano (1981).

Although the sound spectrograph is a useful tool to analyze various acoustic characteristics of voice or speech, quantitative measurements are time consuming. Real-time spectrographic instruments are available today that are based on computers with high-speed digital signal-processing capabilities. Some of those instruments can automatically extract acoustic parameters such as those mentioned in this chapter. Those instruments are convenient for quantitative evaluation of acoustic characteristics of voice to determine the nature and degree of voice disorders and to monitor changes (Gauffin et al, 1986; Hirano, 1989; Imaizumi, 1986; Kasuya et al, 1986).

The sound spectrograms shown in Figs. 96-37 and 96-38 were made using the Kay DIgital Sonagraph MODEL 7800, and those shown in Fig. 96-39 were made using the RION SG-07.