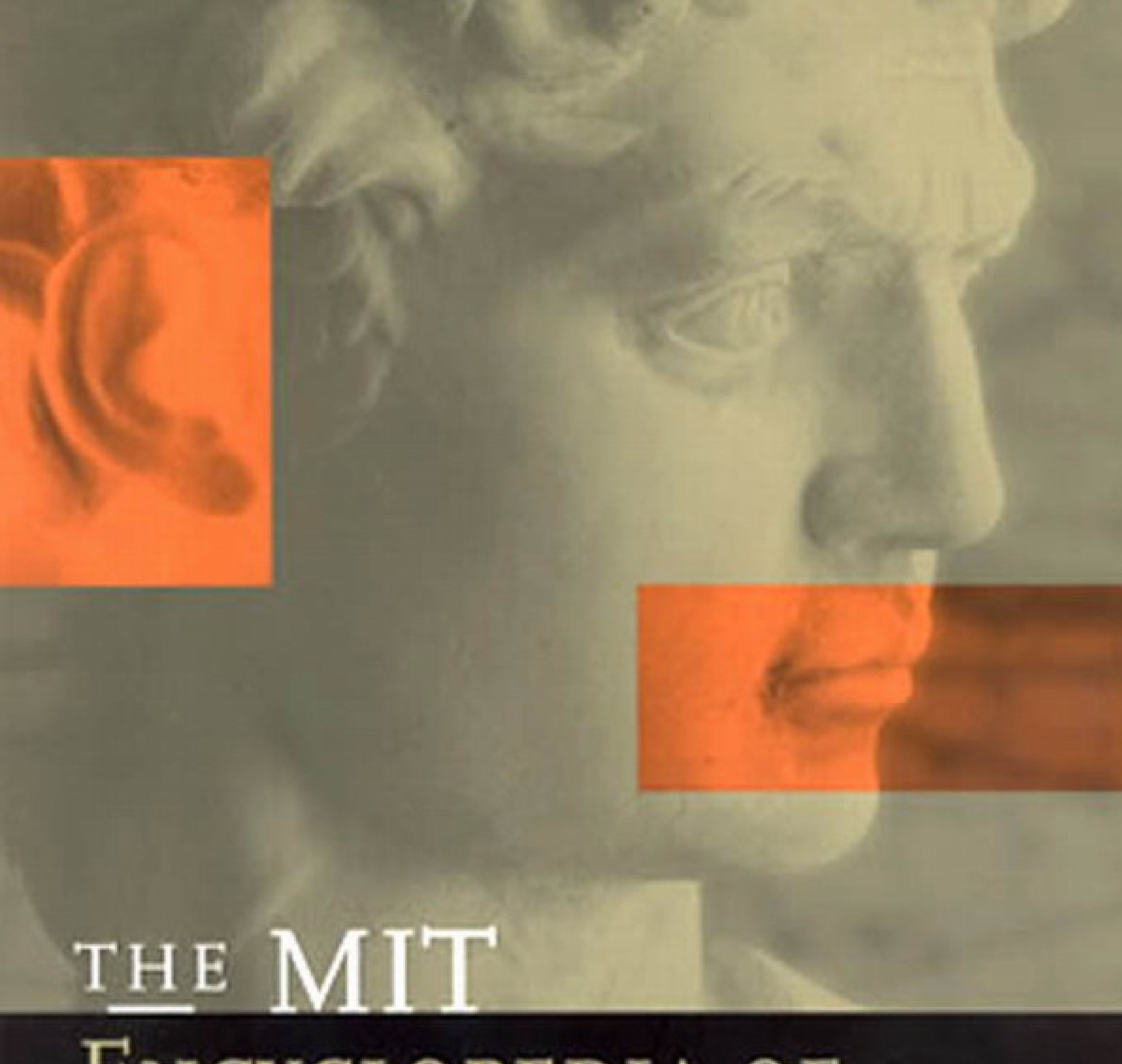

THE MIT ENCYCLOPEDIA OF COMMUNICATION DISORDERS

EDITED BY

RAYMOND D. KENT

The MIT Encyclopedia of Communication Disorders

The MIT Encyclopedia of Communication Disorders

Edited by Raymond D. Kent

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

This book was set in Times New Roman on 3B2 by Asco Typesetters, Hong Kong, and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

The MIT encyclopedia of communication disorders / edited by Raymond D. Kent.
p. cm.
Includes bibliographical references and index.
ISBN 0-262-11278-7 (cloth)
1. Communicative disorders—Encyclopedias. I. Kent, Raymond D. II. Massachusetts Institute of Technology.
RC423.M56 2004
616.85′5′003—dc21

2003059941

Contents

Part III: Language 229

Part IV: Hearing 411

Introduction

The MIT Encyclopedia of Communication Disorders (MITECD) is a comprehensive volume that presents essential information on communication sciences and disorders. The pertinent disorders are those that affect the production and comprehension of spoken language and include especially disorders of speech production and perception, language expression, language comprehension, voice, and hearing. Potential readers include clinical practitioners, students, and research specialists. Relatively few comprehensive books of similar design and purpose exist, so MITECD stands nearly alone as a resource for anyone interested in the broad field of communication disorders.

MITECD is organized into the four broad categories of Voice, Speech, Language, and Hearing. These categories represent the spectrum of topics that usually fall under the rubric of communication disorders (also known as speech-language pathology and audiology, among other names). For example, roughly these same categories were used by the National Institute on Deafness and Other Communication Disorders (NIDCD) in preparing its national strategic research plans over the past decade. The *Journal of Speech, Language, and Hearing Research*, one of the most comprehensive and influential periodicals in the field, uses the editorial categories of speech, language, and hearing. Although voice could be subsumed under speech, the two fields are large enough individually and sufficiently distinct that a separation is warranted. Voice is internationally recognized as a clinical and research specialty, and it is represented by journals dedicated to its domain (e.g., the *Journal of Voice*). The use of these four categories achieves a major categorization of knowledge but avoids a narrow fragmentation of the field at large. It is to be expected that the *Encyclopedia* would include cross-referencing within and across these four major categories. After all, they are integrated in the definitively human behavior of language, and disorders of communication frequently have wide-ranging effects on communication in its essential social, educational, and vocational roles.

In designing the content and structure of MITECD, it was decided that each of these major categories should be further subdivided into Basic Science, Disorders (nature and assessment), and Clinical Management (intervention issues). Although these categories are not always transparent in the entire collection of entries, they guided the delineation of chapters and the selection of contributors. These categories are defined as follows:

Basic Science entries pertain to matters such as normal anatomy and physiology, physics, psychology and psychophysics, and linguistics. These topics are the foundation for clinical description and interpretation, covering basic principles and terminology pertaining to the communication sciences. Care was taken to avoid substantive overlap with previous MIT publications, especially the *MIT Encyclopedia of the Cognitive Sciences* (MITECS).

The Disorders entries offer information on issues such as syndrome delineation, definition and characterization of specific disorders, and methods for the identification and assessment of disorders. As such, these chapters reflect contemporary nosology and nomenclature, as well as guidelines for clinical assessment and diagnosis.

The Clinical Management entries discuss various interventions including behavioral, pharmacological, surgical, and prosthetic (mechanical and electronic). There is a general, but not necessarily one-to-one, correspondence between chapters in the Disorders and Clinical Management categories. For example, it is possible that several types of disorder are related to one general chapter on clinical management. It is certainly the case that different management strategies are preferred by different clinicians. The chapters avoid dogmatic statements regarding interventions of choice.

Because the approach to communicative disorders can be quite different for children and adults, a further cross-cutting division was made such that for many topics separate chapters for children and adults are included. Although some disorders that are first diagnosed in childhood may persist in some form throughout adulthood (e.g, stuttering, specific language impairment, and hearing loss may be lifelong conditions for some individuals), many disorders can have an onset either in childhood or in adulthood and the timing of onset can have implications for both assessment and intervention. For instance, when a child experiences a significant loss of hearing, the sensory deficit may greatly impair the learning of speech and language. But when a loss of the same degree has an onset in adulthood, the problem is not in acquiring speech and language, but rather in maintaining communication skills. Certainly, it is often true that an understanding of a given disorder has common features in both the developmental and acquired forms, but commonality cannot be assumed as a general condition.

Many decisions were made during the preparation of this volume. Some were easy, but others were not. In the main, entries are uniform in length and number of references. However, in a few instances, two or more entries were combined into a single longer entry. Perhaps inevitably in a project with so many contributors, a small number of entries were dropped because of personal issues, such as illness, that interfered with timely preparation of an entry. Happily, contributors showed great enthusiasm for this project, and their entries reflect an assembled expertise that is high tribute to the science and clinical practice in communication disorders.

Raymond D. Kent

Acknowledgments

MITECD began as a promising idea in a conversation with Amy Brand, a previous editor with MIT Press. The idea was further developed, refined, elaborated, and refined again in many ensuing e-mail communications, and I thank Amy for her constant support and assistance through the early phases of the project. When she left MIT Press, Tom Stone, Senior Editor of Cognitive Sciences, Linguistics, and Bradford Books, stepped in to provide timely advice and attention. I also thank Mary Avery, Acquisitions Assistant, for her help in keeping this project on track. I am indebted to all of them.

Speech, voice, language, and hearing are vast domains individually, and several associated editors helped to select topics for inclusion in MITECD and to identify contributors with the necessary expertise. The associate editors and their fields of responsibility are as follows:

Fred H. Bess, Ph.D., Hearing Disorders in Children
Joseph R. Duffy, Ph.D., Speech Disorders in Adults
Steven D. Gray, M.D. (deceased), Voice Disorders in Children
Robert E. Hillman, Ph.D., Voice Disorders in Adults
Sandra Gordon-Salant, Ph.D., Hearing Disorders in Adults
Mabel L. Rice, Ph.D., Language Disorders in Children
Lawrence D. Shriberg, Ph.D., Speech Disorders in Children
David A. Swinney, Ph.D., and Lewis P. Shapiro, Ph.D., Language Disorders in Adults

The advice and cooperation of these individuals is gratefully acknowledged. Sadly, Dr. Steven D. Gray died within the past year. He was an extraordinary man, and although I knew him only briefly, I was deeply impressed by his passion for knowledge and life. He will be remembered as an excellent physician, creative scientist, and valued friend and colleague to many.

Dr. Houri Vorperian greatly facilitated this project through her inspired planning of a computer-based system for contributor communications and record management. Sara Stuntebeck and Sara Brost worked skillfully and accurately on a variety of tasks that went into different phases of MITECD. They offered vital help with communications, file management, proofreading, and the various and sundry tasks that stood between the initial conception of MITECD and the submission of a full manuscript.

P. M. Gordon and Associates took on the formidable task of assembling 200 entries into a volume that looks and reads like an encyclopedia. I thank Denise Bracken for exacting attention to the editing craft, creative solutions to unexpected problems, and forbearance through it all.

MITECD came to reality through the efforts of a large number of contributors—too many for me to acknowledge personally here. However, I draw the reader's attention to the list of contributors included in this volume. I feel a sense of community with all of them, because they believed in the project and worked toward its completion by preparing entries of high quality. I salute them not only for their contributions to MITECD but also for their many career contributions that define them as experts in the field. I am honored by their participation and their patient cooperation with the editorial process.

Raymond D. Kent

Part I: Voice

Acoustic Assessment of Voice

Acoustic assessment of voice in clinical applications is dominated by measures of fundamental frequency (f_0), cycle-to-cycle perturbations of period (jitter) and intensity (shimmer), and other measures of irregularity, such as noise-to-harmonics ratio (NHR). These measures are widely used, in part because of the availability of electronic and microcomputer-based instruments (e.g., Kay Elemetrics Computerized Speech Laboratory [CSL] or Multispeech, Real-Time Pitch, Multi-Dimensional Voice Program [MDVP], and other software/hardware systems), and in part because of long-term precedent for perturbation (Lieberman, 1961) and spectral noise measurements (Yanagihara, 1967). Absolute measures of vocal intensity are equally basic but require calibrations and associated instrumentation (Winholtz and Titze, 1997).

Independently, these basic acoustic descriptors—f_0, intensity, jitter, shimmer, and NHR—can provide some very basic characterizations of vocal health. The first two, f_0 and intensity, have very clear perceptual correlates—pitch and loudness, respectively—and should be assessed for both stability and variability and compared to age and sex norms (Kent, 1994; Baken and Orlikoff, 2000). Ideally, these tasks are recorded over headset microphones with direct digital acquisition at very high sampling rates (at least 48 kHz). The materials to be assessed should be obtained following standardized elicitation protocols that include sustained vowel phonations at habitual levels, levels spanning a client's vocal range in both f_0 and intensity, running speech, and speech tasks designed to elicit variation (Titze, 1995; Awan, 2001). Note, however, that not all measures will be appropriate for all tasks; perturbation statistics, for example, are usually valid only when extracted from sustained vowel phonations.

These basic descriptors are not in any way comprehensive of the range of available measures or the available signal properties and dimensions. Table 1 categorizes measures (Buder, 2000) based on primary basic signal representations from which measures are derived. Although these categories are intended to be exhaustive and mutually exclusive, some more modern algorithms process components through several types. (For more detail on the measurement types, see Buder, 2000, and Baken and Orlikoff, 2000.) Modern algorithmic approaches should be selected for (1) interpretability with respect to aerodynamic and physiological models of phonation and (2) the incorporation of multivariate measures to characterize vocal function.

Interdependence of Basic Measures. The interdependence between f_0 and intensity is mapped in a voice range profile, or phonetogram, which is an especially valuable assessment for the professional voice user (Coleman, 1993). Furthermore, the dependence of perturbations and signal-to-noise ratios on both f_0 and intensity is well known (Klingholz, 1990; Pabon, 1991).

Table 1. Outline of Traditional Acoustic Algorithm Types

f_0 statistics
Short-term perturbations
Long-term perturbations
Amplitude statistics
Short-term perturbations
Long-term perturbations
f_0/amplitude covariations
Waveform perturbations
Spectral measures
Spectrographic measures
Fourier and LPC spectra
Long-term average spectra
Cepstra
Inverse filter measures
Radiated signal
Flow-mask signals
Dynamic measures

This dependence is not often assessed rigorously, perhaps because of the time-consuming and strenuous nature of a full voice profile. However, an abbreviated or focused profiling in which samples related to habitual f_0 by a set number of semitones, or related to habitual intensity by a set number of decibels, could be standardized to control for this dependence efficiently. Finally, it should be understood that perturbations and NHR-type measures will usually covary for many reasons, the simplest ones being methodological (Hillenbrand, 1987): an increase in any one of the underlying phenomena detected by a single measure will also affect the other measures.

Periodicity as a Reference. The chief problem with nearly all acoustic assessments of voice is the determination of f_0. Most voice quality algorithms are based on the prior identification of the periodic component in the signal (based on glottal pulses in the time domain or harmonic structure in the frequency domain). Because phonation is ideally a nearly periodic process, it is logical to conceive of voice measures in terms of the degree to which a given sample deviates from pure periodicity. There are many conceptual problems with this simplification, however. At the physiological level, glottal morphology is multidimensional—superior-inferior asymmetry is a basic feature of the two-mass model (Ishizaka and Flanagan, 1972), and some anterior-posterior asymmetry is also inevitable—rendering it unlikely that a glottal pulse will be marked by a discrete or even a single instant of glottal closure. At the level of the signal, the deviations from periodicity may be either random or correlated, and in many cases they are so extreme as to preclude identification of a regular period. Finally, at the perceptual level, many factors related to deviations from a pure f_0 can contribute to pitch perception (Zwicker and Fastl, 1990).

At any or all of these levels, it becomes questionable to characterize deviations with pure periodicity as a reference. In acoustic assessment, the primary level of concern is the signal. The National Center for Voice and

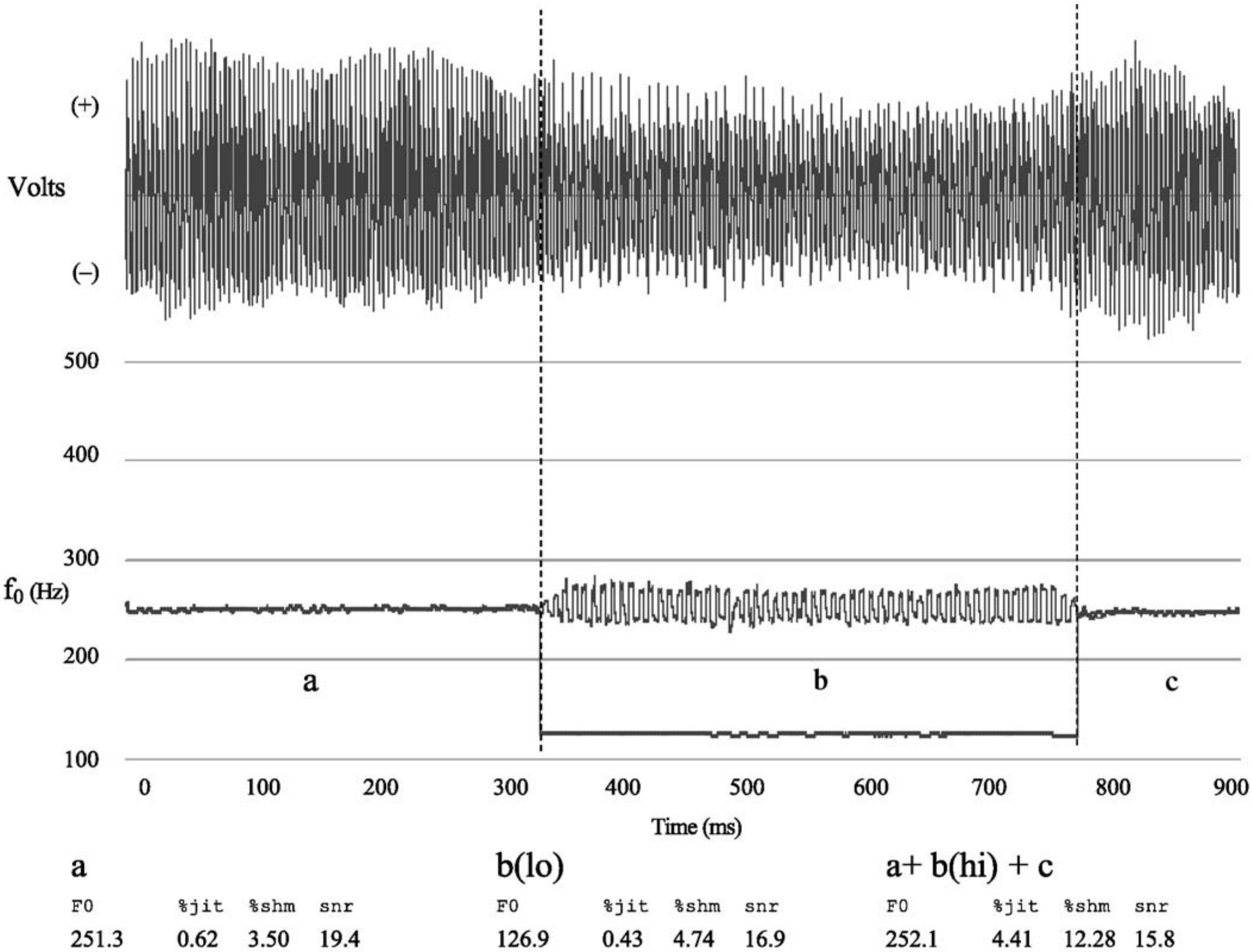

Figure 1. Approximately 900 ms of a sustained vowel phonation waveform (top panel) with two fundamental frequency analyses (bottom panel). Average f_0, %jitter, %shimmer, and SNR results for selected segments were from the "newjit" routine of TF32 program (Milenkovic, 2001).

Speech issued a summary statement (Titze, 1995) recommending a typology for categorizing deviations from periodicity in voices (see also Baken and Orlikoff, 2000, for further subtypes). This typology capitalizes on the categorical nature of dynamic states in nonlinear systems; all the major categories, including stable points, limit cycles, period-doubling/tripling/. . . , and chaos can be observed in voice signals (Herzel et al., 1994; Sataloff and Hawkshaw, 2001). As in most highly nonlinear dynamic systems, deviations from periodicity can be categorized on the basis of bifurcations, or sudden qualitative changes in vibratory pattern from one of these states to another.

Figure 1 displays a common form for one such bifurcation and illustrates the importance of accounting for its presence in the application of perturbation measures. In this sustained vowel phonation by a middle-aged woman with spasmodic dysphonia, a transition to subharmonics is clearly visible in segment *b* (similar patterns occur in individuals without dysphonias). Two f_0 extractions are presented for this segment, one at the targeted level of approximately 250 Hz and another which the tracker finds one octave below this; inspection of the waveform and a perceived biphonia both justify this 125-Hz analysis as a new fundamental frequency, although it can also be understood in this context as a subharmonic to the original fundamental. There is therefore some ambiguity as to which fundamental is valid during this episode, and an automatic analysis could plausibly identify either frequency. (Here the waveform-matching algorithm implemented in CSpeechSP [Milenkovic, 1997] does identify either frequency, depending on where in the waveform the algorithm is applied; initiating the algorithm within the subharmonic segment predisposes it to identify the lower fundamental.)

The acoustic measures of the segments displayed in Figure 1 reveal the nontrivial differences that result, depending on the basic glottal pulse form under consideration. When the pulses of segment *a* are considered, the perturbations around the base period associated with the high f_0 are low and normative; in segment *b*, perturbations around the longer periods of the lower f_0 are still low (jitter is improved, while shimmer and the signal-to-noise ratio show some degradation). However, when all segments are considered together to include the perturbations around the high f_0 tracked through segment *b* and into *c*, the perturbation statistics are all increased by an order of magnitude. Many important methodological and theoretical questions should be raised by such common scenarios in which we must consider not just voice typing, but the segment-by-segment validity of applying perturbation measures with a particular f_0 as reference. If, as is often assumed, jitter and shimmer are ascribed to ''random'' variations, then the correlated modulations of a strong subharmonic episode should be excluded. Alternatively, the perturbations might be analyzed with respect to the subharmonic f_0. In any case, assessment by means of perturbation statistics with no consideration of their underlying sources is unwise.

Perceptual, Aerodynamic, and Physiological Correlates of Acoustic Measures. Regarding perceptual voice ratings, Gerratt and Kreiman (2000) have critiqued traditional assessments on several important methodological and theoretical points. However, these points may not apply to acoustic analysis if (1) acoustic analysis is validated on its own success and not exclusively in relation to the problematic perceptual classifications, and (2) acoustic analysis is thoroughly grounded for interpretation in some clear aerodynamic or physiological model of phonation. Gerratt and Kreiman also argue that clinical classification may not be derived along a continuum that is defined with reference to normal qualities, but again, this argument may need to be reversed for the acoustic domain. It is only by reference to a specific model that any assessment on acoustic grounds can be interpreted (though this does not preclude development of an independent model for a pathological phonatory mechanism). In clinical settings, acoustic voice assessment often serves to corroborate perceptual assessment. However, as guided by auditory experience and in conjunction with the ear and other instrumental assessments, careful acoustic analysis can be oriented to the identification of physiological status.

In attempting to draw safe and reasonably direct inferences from acoustic signal, aerodynamic models

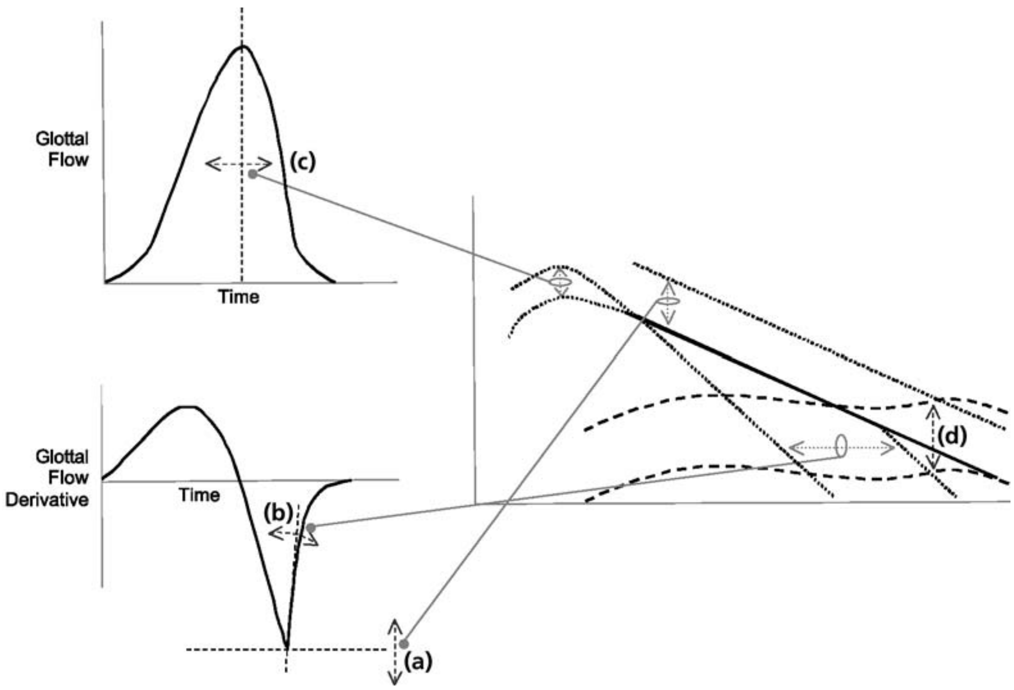

Figure 2. Spectral features associated with models of phonation, including the Liljencrants-Fant (LF) model of glottal flow and aperiodicity source models developed by Stevens. The LF model of glottal flow is shown at top left. At bottom left is the LF model of glottal flow derivative, showing the rate of change in flow. At right is a spectrum schematic showing four effects. These effects include three derived parameters of the LF model: (**a**) excitation strength (the maximum negative amplitude of the flow derivative, which is positively correlated with overall harmonic energy), (**b**) dynamic leakage or non-zero return phase following the point of maximum excitation (which is negatively correlated with high-frequency harmonic energy), and (**c**) pulse skewing (which is negatively correlated with low-frequency harmonic energy; this low-frequency region is also positively correlated with open quotient and peak volume velocity measures of the glottal flow waveform). The effect of turbulence due to high airflow through the glottis is schematized by (**d**), indicating the associated appearance of high-frequency aperiodic energy in the spectrum. See VOICE ACOUSTICS for other graphical and quantitative associations between glottal status and spectral characteristics.

of glottal behavior present important links to the physiological domain. Attempts to recover the glottal flow waveform, either from a face mask-transduced flow recording (Rothenberg, 1973) or a microphone-transduced acoustic recording (Davis, 1975), have proved to be labor-intensive and prone to error (Ní Chasaide and Gobl, 1997). Rather than attempting to eliminate the effects of the vocal tract, it may be more fruitful to understand its in situ relationship with phonation, and infer, via the types of features displayed in Figure 2, the status of the glottis as a sound source. Interpretation of spectral features, such as the amplitudes of the first harmonics and at the formant frequencies, may be an effective alternative when guided by knowledge of glottal aerodynamics and acoustics (Hanson, 1997; Ní Chasaide and Gobl, 1997; Hanson and Chuang, 1999). Deep familiarity with acoustic mechanisms is essential for such interpretations (Titze, 1994; Stevens, 1998), as is a model with clear and meaningful parameters, such as the Liljencrants-Fant (LF) model (Fant, Liljencrants, and Lin, 1985). The parameters of the LF model have proved to be meaningful in acoustic studies (Gauffin & Sundberg, 1989) and useful in refined efforts at inverse filtering (Fröhlich, Michaelis, and Strube, 2001). Figure 2 summarizes selected parameters of the LF source model following Ní Chasaide and Gobl (1997) and the glottal turbulence source following Stevens (1998); see also VOICE ACOUSTICS for other approaches relating glottal status to spectral measures.

Other spectral-based measures implement similar model-based strategies by selecting spectral component ratios (e.g., the VTI and SPI parameters of MDVP). Sophisticated spectral noise characterizations control for perturbations and modulations (Murphy, 1999; Qi, Hillman, and Milstein, 1999), or employ curve-fitting and statistical models to produce more robust measures (Alku, Strik, and Vilkman, 1997; Michaelis, Fröhlich, and Strube, 1998; Schoentgen, Bensaid, and Bucella, 2000). A particularly valuable modern technique for detecting turbulence at the glottis, the glottal-to-noise-excitation ratio (Michaelis, Gramss, and Strube, 1997), has been especially successful in combination with other measures (Fröhlich et al., 2000). The use of acoustic techniques for voice will only improve with the inclusion of more knowledge-based measures in multivariate representations (Wolfe, Cornell, and Palmer, 1991; Callen et al., 2000; Wuyts et al., 2000).

—*Eugene H. Buder*

References

Alku, P., Strik, H., and Vilkman, E. (1997). Parabolic spectral parameter: A new method for quantification of the glottal flow. *Speech Communication*, *22*, 67–79.

Awan, S. N. (2001). *The voice diagnostic profile: A practical guide to the diagnosis of voice disorders*. Gaithersburg, MD: Aspen.

Baken, R. J., and Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*. San Diego, CA: Singular Publishing Group.

Buder, E. H. (2000). Acoustic analysis of voice quality: A tabulation of algorithms 1902–1990. In M. J. Ball (Ed.), *Voice quality measurement* (pp. 119–244). San Diego, CA: Singular Publishing Group.

Callen, D. E., Kent, R. D., Roy, N., and Tasko, S. M. (2000). The use of self-organizing maps for the classification of voice disorders. In M. J. Ball (Ed.), *Voice quality measurement* (pp. 103–116). San Diego, CA: Singular Publishing Group.

Coleman, R. F. (1993). Sources of variation in phonetograms. *Journal of Voice*, *7*, 1–14.

Davis, S. B. (1975). Preliminary results using inverse filtering of speech for automatic evaluation of laryngeal pathology. *Journal of the Acoustical Society of America*, *58*, SIII.

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

Fröhlich, M., Michaelis, D., and Strube, H. (2001). SIM-simultaneous inverse filtering and matching of a glottal flow model for acoustic speech signals. *Journal of the Acoustical Society of America*, *110*, 479–488.

Fröhlich, M., Michaelis, D., Strube, H., and Kruse, E. (2000). Acoustic voice analysis by means of the hoarseness diagram. *Journal of Speech, Language, and Hearing Research*, *43*, 706–720.

Gauffin, J., and Sundberg, J. (1989). Spectral correlates of glottal voice source waveform characteristics. *Journal of Speech and Hearing Research*, *32*, 556–565.

Gerratt, B., and Kreiman, J. (2000). Theoretical and methodological development in the study of pathological voice quality. *Journal of Phonetics*, *28*, 335–342.

Hanson, H. M. (1997). Glottal characteristics of female speakers: Acoustic correlates. *Journal of the Acoustical Society of America*, *101*, 466–481.

Hanson, H. M., and Chuang, E. S. (1999). Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *Journal of the Acoustical Society of America*, *106*, 1064–1077.

Herzel, H., Berry, D., Titze, I. R., and Saleh, M. (1994). Analysis of vocal disorders with methods from nonlinear dynamics. *Journal of Speech and Hearing Research*, *37*, 1008–1019.

Hillenbrand, J. (1987). A methodological study of perturbation and additive noise in synthetically generated voice signals. *Journal of Speech and Hearing Research*, *30*, 448–461.

Ishizaka, K., and Flanagan, J. L. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell System Technical Journal*, *51*, 1233–1268.

Kent, R. D. (1994). *Reference manual for communicative sciences and disorders: Speech and language*. Austin, TX: Pro-Ed.

Klingholz, F. (1990). Acoustic representation of speaking-voice quality. *Journal of Voice*, *4*, 213–219.

Lieberman, P. (1961). Perturbations in vocal pitch. *Journal of the Acoustical Society of America*, *33*, 597–603.

Michaelis, D., Fröhlich, M., and Strube, H. W. (1998). Selection and combination of acoustic features for the description of pathologic voices. *Journal of the Acoustical Society of America*, *103*, 1628–1638.

Michaelis, D., Gramss, T., and Strube, H. W. (1997). Glottal to noise excitation ratio: A new measure for describing pathololcial voices. *Acustica*, *83*, 700–706.

Milenkovic, P. (1997). CSpeechSP [Computer software]. Madison, WI: University of Wisconsin–Madison.

Milenkovic, P. (2001). TF32 [Computer software]. Madison, WI: University of Wisconsin–Madison.

Murphy, P. J. (1999). Perturbation-free measurement of the harmonics-to-noise ratio in voice signals using pitch synchronous harmonic analysis. *Journal of the Acoustical Society of America*, *105*, 2866–2881.

Ní Chasaide, A., and Gobl, C. (1997). Voice source variation. In J. Laver (Ed.), *The handbook of phonetic sciences* (pp. 427–461). Oxford, UK: Blackwell.

Pabon, J. P. H. (1991). Objective acoustic voice-quality parameters in the computer phonetogram. *Journal of Voice*, *5*, 203–216.

Qi, Y., Hillman, R., and Milstein, C. (1999). The estimation of signal-to-noise ratio in continuous speech for disordered voices. *Journal of the Acoustical Society of America*, *105*, 2532–2535.

Rothenberg, M. (1973). A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *Journal of the Acoustical Society of America*, *53*, 1632–1645.

Sataloff, R. T., and Hawkshaw, M. (Eds.). (2001). *Chaos in medicine: Source readings*. San Diego, CA: Singular Publishing Group.

Schoentgen, J., Bensaid, M., and Bucella, F. (2000). Multivariate statistical analysis of flat vowel spectra with a view to characterizing dysphonic voices. *Journal of Speech, Language, and Hearing Research*, *43*, 1493–1508.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.

Titze, I. R. (1995). *Workshop on acoustic voice analysis: Summary statement*. Iowa City, IA: National Center for Voice and Speech.

Winholtz, W. S., and Titze, I. R. (1997). Conversion of a head-mounted microphone signal into calibrated SPL units. *Journal of Voice*, *11*, 417–421.

Wolfe, V., Cornell, R., and Palmer, C. (1991). Acoustic correlates of pathologic voice types. *Journal of Speech and Hearing Research*, *34*, 509–516.

Wuyts, F. L., De Bodt, M. S., Molenberghs, G., Remacle, M., Heylen, L., Millet, B., et al. (2000). The Dysphonia Severity Index: An objective measure of vocal quality based on a multiparameter approach. *Journal of Speech, Language, and Hearing Research*, *43*, 796–809.

Yanagihara, N. (1967). Significance of harmonic changes and noise components in hoarseness. *Journal of Speech and Hearing Research*, *10*, 531–541.

Zwicker, E., and Fastl, H. (1990). *Psychoacoustics: Facts and models*. Heidelberg, Germany: Springer-Verlag.

Aerodynamic Assessment of Vocal Function

A number of methods have been used to quantitatively assess the air volumes, airflows, and air pressures involved in voice production. The methods have been mostly used in research to investigate mechanisms that underlie normal and disordered voice and speech production. The clinical use of aerodynamic measures to assess patients with voice disorders has been increasing (Colton and Casper, 1996; Hillman, Montgomery, and Zeitels, 1997; Hillman and Kobler, 2000).

Measurement of Air Volumes. Respiratory research in human communication has focused primarily on the measurement of the air volumes that are typically expended during selected speech and singing tasks, and on specifying the ranges of lung inflation levels across which such tasks are normally performed (cf. Hixon, Goldman, and Mead, 1973; Watson and Hixon, 1985; Hoit and Hixon, 1987; Hoit et al., 1990). Air volumes are measured in standard metric units (liters, cubic centimeters, milliliters) and lung inflation levels are usually specified in terms of a percentage of the vital capacity or total lung volume.

Both direct and indirect methods have been used to measure air volumes expended during phonation. Direct measurement of orally displaced air volumes during phonatory tasks can be accomplished, to a limited extent, by means of a mouthpiece or face mask connected to a measurement device such as a spirometer (Beckett, 1971) or pneumotachograph (Isshiki, 1964). The use of a mouthpiece essentially limits speech production to sustained vowels, which are sufficient for assessing selected volumetric-based phonatory parameters. There are also concerns that face masks interfere with normal jaw movements and that the oral acoustic signal is degraded, so that auditory feedback is reduced or distorted and simultaneous acoustic analysis is limited. These limitations, which are inherent to the use of devices placed in or around the mouth to directly collect oral airflow, plus additional measurement-related restrictions (Hillman and Kobler, 2000) have helped motivate the development and application of indirect measurement approaches.

Most speech breathing research has been carried out using indirect approaches for estimating lung volumes by means of monitoring changes in body dimensions. The basic assumption underlying the indirect approaches is that changes in lung volume are reflected in proportional changes in body torso size. One relatively cumbersome but time-honored approach has been to place subjects in a sealed chamber called a body plethysmograph to allow estimation of the air volume displaced by the body during respiration (Draper, Ladefoged, and Whitteridge, 1959). More often used for speech breathing research are transducers (magnetometers: Hixon, Goldman, and Mead, 1973; inductance plethysmographs: Sperry, Hillman, and Perkell, 1994) that unobtrusively monitor changes in the dimensions of the rib cage and abdomen (referred to collectively as the chest wall) that account for the majority of respiratory-related changes in torso dimension (Mead et al., 1967). These approaches have been primarily employed to study respiratory function during continuous speech and singing tasks that include both voiced and voiceless sound production, as opposed to assessing air volume usage during phonatory tasks that involve only laryngeal production of voice (e.g., sustained vowels). There are also ongoing efforts to develop more accurate methods for noninvasively monitoring chest wall activity to capture finer details of how the three-dimensional geometry of the body is altered during respiration (see Cala et al., 1996).

Measurement of Airflow. Airflow associated with phonation is usually specified in terms of volume velocity (i.e., volume of air displaced per unit of time). Volume velocity airflow rates for voice production are typically reported in metric units of volume displaced (liters or cubic centimeters) per second.

Estimates of average airflow rates can be obtained by simply dividing air volume estimates by the duration of the phonatory task. Average glottal airflow rates have usually been estimated during vowel phonation by using a mouthpiece or face mask to channel the oral air stream through a pneumotachograph (Isshiki, 1964). There has also been somewhat limited use of hot wire anemometer devices (mounted in a mouthpiece) to estimate average glottal airflow during sustained vowel phonation (Woo, Colton, and Shangold, 1987). Estimates of average glottal airflow rates can be obtained from the oral airflow during vowel production because the vocal tract is relatively nonconstricted, with no major sources of turbulent airflow between the glottis and the lips.

There have also been efforts to obtain estimates of the actual airflow waveform that is generated as the glottis rapidly opens and closes during flow-induced vibration of the vocal folds (the glottal volume velocity waveform). The glottal volume velocity waveform cannot be directly observed by measuring the oral airflow signal because the waveform is highly convoluted by the resonance activity (formants) of the vocal tract. Thus, recovery of the glottal volume velocity waveform requires methods that eliminate or correct for the influences of the vocal tract. This has typically been accomplished aerodynamically by processing the output of a fast-responding pneumotachograph (high-frequency response) using a technique called inverse filtering, in which the major resonances of the vocal tract are estimated and the oral airflow signal is processed (inverse filtered) to eliminate them (Rothenberg, 1977; Holmberg, Hillman, and Perkell, 1988).

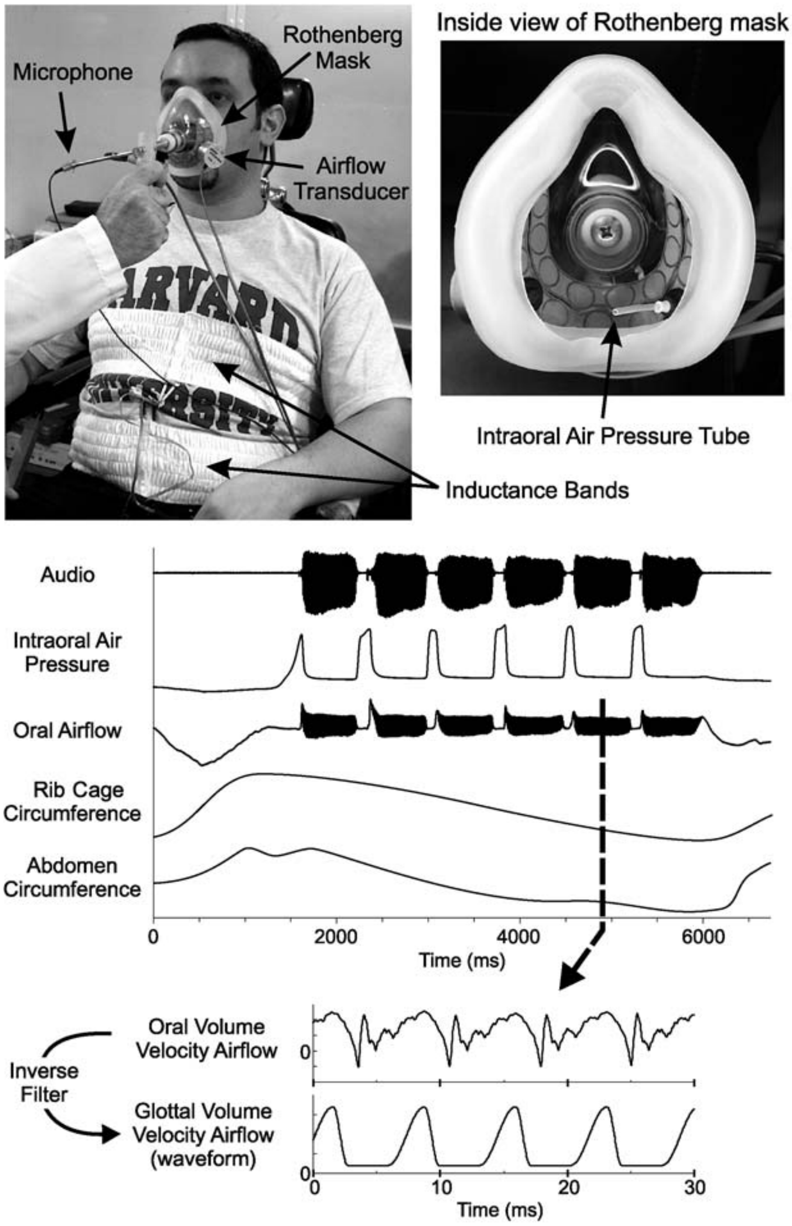

Figure 1. Instrumentation and resulting signals for simultaneous collection of oral airflow, intraoral air pressure, the acoustic signal, and chest wall (rib cage and abdomen) dimensions during production of the syllable string /pi-pi-pi/. Signals shown in the bottom panel are processed and measured to provide estimates of average glottal airflow rate, average subglottal air pressure, lung volume, and glottal waveform parameters.

Measurement of Air Pressure. Measurements of air pressures below (subglottal) and above (supraglottal) the vocal folds are of primary interest for characterizing the pressure differential that must be achieved to initiate and maintain vocal fold vibration during normal exhalatory phonation. In practice, air pressure measurements related specifically to voice production are typically acquired during vowel phonation when there are no vocal tract constrictions of sufficient magnitude to build up positive supraglottal pressures. Under these conditions, it is usually assumed that supraglottal pressure is essentially equal to atmospheric pressure and only subglottal pressure measurements are obtained. Air pressures associated with voice and speech production are usually specified in centimeters of water (cm H_2O).

Both direct and indirect methods have been used to measure subglottal air pressures during phonation. Direct measures of subglottal air pressure can be obtained by inserting a hypodermic needle into the subglottal airway through a puncture in the anterior neck at the cricothyroid space (Isshiki, 1964). The needle is connected to a pressure transducer by tubing. This method is very accurate but also very invasive. It is also possible to insert a very thin catheter through the posterior cartilaginous glottis (between the arytenoids) to sense subglottal air pressure during phonation, or to use an array of miniature transducers positioned directly above and below the glottis (Cranen and Boves, 1985). These methods cannot be tolerated by all subjects, and the heavy topical anesthetization of the larynx that is required can affect normal function.

Indirect estimates of tracheal (subglottal) air pressure can be obtained via the placement of an elongated balloon-like device into the esophagus (Liberman, 1968). The deflated esophageal balloon is attached to a catheter that is typically inserted transnasally and then swallowed into the esophagus to be positioned at the midthoracic level. The catheter is connected to a pressure transducer and the balloon is slightly inflated. Accurate use of this invasive method also requires simultaneous monitoring of lung volume.

Noninvasive, indirect estimates of subglottal air pressure can be obtained by measuring intraoral air pressure during specially constrained utterances (Smitheran and Hixon, 1981). This is usually done by sensing air pressure just behind the lips with a translabially placed catheter connected to a pressure transducer. These intraoral pressure measures are obtained as subjects produce strings of bilabial /p/ + vowel syllables (e.g., /pi-pi-pi-pi-pi/) at constant pitch and loudness. This method works because the vocal folds are abducted during /p/ production, thus allowing pressure to equilibrate throughout the airway, making intraoral pressure equal to subglottal pressure (Fig. 1).

Additional Derived Measures. There have been numerous attempts to extend the utility of aerodynamic measures by using them in the derivation of additional parameters aimed at better elucidating underlying mechanisms of vocal function. Such derived measures usually take the form of ratios that relate aerodynamic parameters to each other, or that relate aerodynamic parameters to simultaneously obtained acoustic measures. Common examples include (1) airway (glottal) resistance (see Smitheran and Hixon, 1981), (2) vocal efficiency (Schutte, 1980; Holmberg, Hillman, and Perkell, 1988), and (3) measures that interrelate glottal volume velocity waveform parameters (Holmberg, Hillman, and Perkell, 1988).

Normative Data. As is the case for most measures of vocal function, there is not currently a set of normative data for aerodynamic measures that is universally accepted and applied in research and clinical work. Methods for collecting such data have not been standardized, and study samples have generally not been of sufficient size or appropriately stratified in terms of age and sex to ensure unbiased estimates of underlying aerodynamic phonatory parameters in the normal population. However, there are several sources in the literature that provide estimates of normative values for selected aerodynamic measures (Kent, 1994; Baken, 1996; Colton and Casper, 1996).

See also VOICE PRODUCTION: PHYSICS AND PHYSIOLOGY.

—*Robert E. Hillman*

References

Baken, R. J. (1996). *Clinical measurement of voice and speech*. San Diego, CA: Singular Publishing Group.

Beckett, R. L. (1971). The respirometer as a diagnostic and clinical tool in the speech clinic. *Journal of Speech and Hearing Disorders*, *36*, 235–241.

Cala, S. J., Kenyon, C. M., Ferrigno, G., Carnevali, P., Aliverti, A., Pedotti, A., et al. (1996). Chest wall and lung volume estimation by optical reflectance motion analysis. *Journal of Applied Physiology*, *81*, 2680–2689.

Colton, R. H., and Casper, J. K. (1996). *Understanding voice problems: A physiological perspective for diagnosis and treatment*. Baltimore: Williams and Wilkins.

Cranen, B., and Boves, L. (1985). Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models for speech production. *Journal of the Acoustical Society of America*, *77*, 1543–1551.

Draper, M., Ladefoged, P., and Whitteridge, P. (1959). Respiratory muscles in speech. *Journal of Speech and Hearing Research*, *2*, 16–27.

Hillman, R. E., and Kobler, J. B. (2000). Aerodynamic measures of voice production. In R. Kent and M. Ball (Eds.), *The handbook of voice quality measurement*, San Diego, CA: Singular Publishing Group.

Hillman, R. E., Montgomery, W. M., and Zeitels, S. M. (1997). Current diagnostics and office practice: Use of objective measures of vocal function in the multidisciplinary management of voice disorders. *Current Opinion in Otolaryngology–Head and Neck Surgery*, *5*, 172–175.

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

Hoit, J. D., and Hixon, T. J. (1987). Age and speech breathing. *Journal of Speech and Hearing Research*, *30*, 351–366.

Hoit, J. D., Hixon, T. J., Watson, P. J., and Morgan, W. J. (1990). Speech breathing in children and adolescents. *Journal of Speech and Hearing Research*, *33*, 51–69.

Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (1988). Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice [published erratum appears in *Journal of the Acoustical Society of America*, 1989, *85*(4), 1787]. *Journal of the Acoustical Society of America*, *84*, 511–529.

Isshiki, N. (1964). Regulatory mechanisms of vocal intensity variation. *Journal of Speech and Hearing Research*, *7*, 17–29.

Kent, R. D. (1994). *Reference manual for communicative sciences and disorders*. San Diego, CA: Singular Publishing Group.

Lieberman, P. (1968). Direct comparison of subglottal and esophageal pressure during speech. *Journal of the Acoustical Society of America*, *43*, 1157–1164.

Mead, J., Peterson, N., Grimgy, N., and Mead, J. (1967). Pulmonary ventilation measured from body surface movements. *Science*, *156*, 1383–1384.

Rothenberg, M. (1977). Measurement of airflow in speech. *Journal of Speech and Hearing Research*, *20*, 155–176.

Schutte, H. (1980). *The efficiency of voice production*. Groningen, The Netherlands: Kemper.

Smitheran, J. R., and Hixon, T. J. (1981). A clinical method for estimating laryngeal airway resistance during vowel production. *Journal of Speech and Hearing Disorders*, *46*, 138–146.

Sperry, E., Hillman, R. E., and Perkell, J. S. (1994). The use of an inductance plethysmograph to assess respiratory function in a patient with nodules. *Journal of Medical Speech-Language Pathology*, *2*, 137–145.

Watson, P. J., and Hixon, T. J. (1985). Respiratory kinematics in classical (opera) singers. *Journal of Speech and Hearing Research*, *28*, 104–122.

Woo, P., Colton, R. H., and Shangold, L. (1987). Phonatory airflow analysis in patients with laryngeal disease. *Annals of Otology, Rhinology, and Laryngology*, *96*, 549–555.

Alaryngeal Voice and Speech Rehabilitation

Loss of the larynx due to disease or injury will result in numerous and significant changes that cross anatomical, physiological, psychological, social, psychosocial, and communication domains. Surgical removal of the larynx, or total laryngectomy, involves resectioning the entire framework of the larynx. Although total laryngectomy may occur in some instances due to traumatic injury, the majority of cases worldwide are the result of cancer. Approximately 75% of all laryngeal tumors arise from squamous epithelial tissue of the true vocal fold (Bailey, 1985). In some instances, and because of the location of many of these lesions, less aggressive approaches to medical intervention may be pursued. This may include radiation therapy or partial surgical resection, which seeks to conserve portions of the larynx, or the use of combined chemoradiation protocols (Hillman et al., 1998; Orlikoff et al., 1999). However, when malignant lesions are sufficiently large or when the location of the tumor threatens the lymphatic compartment of the larynx, total laryngectomy is often indicated for reasons of oncological safety (Doyle, 1994).

Effects of Total Laryngectomy

The two most prominent effects of total laryngectomy as a surgical procedure are change of the normal airway and loss of the normal voicing mechanism for verbal communication. Once the larynx is surgically removed from the top of the trachea, the trachea is brought forward to the anterior midline neck and sutured into place near the sternal notch. Thus, total laryngectomy necessitates that the airway be permanently separated from the upper aerodynamic (oral and pharyngeal) pathway. When the laryngectomy is completed, the tracheal airway will remain separate from the oral cavity, pharynx, and esophagus. Under these circumstances, not only is the primary structure for voice generation lost, but the intimate relationship between the pulmonary system and that of the structures of the upper airway, and consequently the vocal tract, is disrupted. Therefore, if verbal communication is to be acquired and used postlaryngectomy, an alternative method of creating an alaryngeal voice source must be achieved.

Methods of Postlaryngectomy Communication

Following laryngectomy, the most significant communicative component to be addressed via voice and speech rehabilitation is the lost voice source. Once the larynx is removed, some alternative method of providing a new, "alaryngeal" sound source is required. There are two general categories in which an alternative, alaryngeal voice source may be achieved. These categories are best described as intrinsic and extrinsic methods. The distinction between these two methods is contingent on the manner in which the alaryngeal voice source is achieved. Intrinsic alaryngeal methods imply that the alaryngeal voice source is found within the system; that is, alternative physical-anatomical structures are used to generate sound. In contrast, extrinsic methods of alaryngeal speech rely on the use of an external sound source, typically an electronic source, or what is termed the artificial larynx, or the electrolarynx. The fundamental differences between intrinsic and extrinsic methods of alaryngeal speech are discussed below.

Intrinsic Methods of Alaryngeal Speech

The two most prominent methods of intrinsic alaryngeal speech are esophageal speech (Diedrich, 1966; Doyle, 1994) and tracheoesophageal (TE) speech (Singer and Blom, 1980). While these two intrinsic methods of alaryngeal speech are dissimilar in some respects, both rely on generation of an alaryngeal voice source by creating oscillation of tissues in the area of the lower pharynx and upper esophagus. This vibratory structure is somewhat variable in regard to width, height, and location (Diedrich and Youngstrom, 1966; Damste, 1986); hence, the preferred term for this alaryngeal voicing source is the pharyngoesophageal (PE) segment. One muscle that comprises the PE segment is the cricopharyngeal muscle. Beyond the commonality in the use of the PE segment as a vicarious voicing source for both esophageal and TE methods of alaryngeal speech, the manner in which these methods are achieved does differ.

Esophageal Speech. For esophageal speech, the speaker must move air from the oral cavity across the tonically closed PE segment in order to insufflate the esophageal reservoir (located inferior to the PE segment). Two methods of insufflation may be utilized. These methods might be best described as being either direct or indirect approaches to insufflation. Direct methods require the individual speaker to actively manipulate air in the oral cavity to effect a change in pressure. When pressure build-up is achieved in the oral cavity via compression maneuvers, and when the pressure becomes of sufficient magnitude to overcome the muscular resistance of the PE segment, air will move across the segment (inferiorly) into the esophagus. This may be accomplished with nonspeech tasks (tongue maneuvers) or as a result of producing specific sounds (e.g., stop consonants).

In contrast, for the indirect (inhalation) method of air insufflation, the speaker indirectly creates a negative pressure in the esophageal reservoir via rapid inhalation through the tracheostoma. This results in a negative pressure in the esophagus relative to the normal atmospheric pressure within the oral cavity/vocal tract (Diedrich and Youngstrom, 1966; Diedrich, 1968; Doyle, 1994). Air then moves passively across the PE segment in order to equalize pressures between the pharynx and esophagus. Once insufflation occurs, this air can be used to generate PE segment vibration in the same manner following other methods of air insufflation. While a distinction between direct and indirect methods permits increased understanding of the physical requirements for esophageal voice production, many esophageal speakers who exhibit high levels of proficiency will often utilize both methods for insufflation. Regardless of which method of air insufflation is used, this air can then be forced back up across the PE segment, and as a result, the tissue of this sphincter will oscillate. This esophageal sound source can then be manipulated in the upper regions of the vocal tract into the sounds of speech.

The acquisition of esophageal speech is a complex process of skill building that must be achieved under the direction of an experienced instructor. Clinical emphasis typically involves tasks that address four skills believed to be fundamental to functional esophageal speech (Berlin, 1963): (1) the ability to phonate reliably on demand, (2) the ability to maintain a short latency between air insufflation and esophageal phonation, (3) the ability to maintain adequate duration of voicing, and (4) the ability to sustain voicing while articulating. These foundation skills have been shown to reflect those progressive abilities that have historically defined speech skills of "superior" esophageal speakers (Wepman et al., 1953; Snidecor, 1968). However, the successful acquisition of esophageal speech may be limited, for many reasons.

Regardless of which method of insufflation is used, esophageal speakers will exhibit limitations in the physical dimensions of speech. Specifically, fundamental frequency is reduced by about one octave (Curry and Snidecor, 1961), intensity is reduced by about 10 dB SPL from that of the normal speaker (Weinberg, Horii, and Smith, 1980), and the durational characteristics of speech are also reduced. Speech intelligibility is also decreased due to limits in the aerodynamic and voicing characteristics of esophageal speech. As it is not an abductory-adductory system, voiced-for-voiceless perceptual errors (e.g., perceptual identification of *b* for *p*) are common. This is a direct consequence of the esophageal speaker's inability to insufflate large or continuous volumes of air into the reservoir. Esophageal speakers must frequently reinsufflate the esophageal reservoir to maintain voicing. Because of this, it is not uncommon to see esophageal speakers exhibit pauses at unusual points in an utterance, which ultimately alters the normal rhythm of speech. Similarly, the prosodic contour of esophageal speech and associated features is often perceived to be abnormal. In contrast to esophageal speech, the TE method capitalizes on the individual's access to pulmonary air for esophageal insufflation, which offers several distinct advantages relative to esophageal speech.

Tracheoesophageal Speech. TE speech uses the same voicing source as traditional esophageal speech, the PE segment. However, in TE speech the speaker is able to access and use pulmonary air as a driving source. This is achieved by the surgical creation of a controlled midline puncture in the trachea, followed by insertion of a one-way TE puncture voice prosthesis (Singer and Blom, 1980), either at the time of laryngectomy or as a second procedure at some point following laryngectomy. Thus, TE speech is best described as a surgical-prosthetic method of voice restoration. Though widely used, TE voice restoration is not problem-free. Limitations in application must be considered, and complications may occur.

The design of the TE puncture voice prosthesis is such that when the tracheostoma is occluded, either by hand or via use of a complementary tracheostoma breathing valve, air is directed from the trachea through the prosthesis and into the esophageal reservoir. This access permits a variety of frequency, intensity, and durational variables to be altered in a fashion different from that of the traditional esophageal speaker (Robbins et al., 1984; Pauloski, 1998). Because the TE speaker has direct access to a pulmonary air source, his or her ability to modify the physical (frequency, intensity, and durational) characteristics of the signal in response to changes in the aerodynamic driving source, along with associated changes in prosodic elements of the speech signal (i.e., stress, intonation, juncture), is enhanced considerably. Such changes have a positive impact on auditory-perceptual judgments of this method of alaryngeal speech.

While the frequency of TE speech is still reduced from that of normal speech, the intensity is greater, and the