ELSEVIER

# ACOUSTIC STUDIES OF DYSARTHRIC SPEECH: METHODS, PROGRESS, AND POTENTIAL

RAY D. KENT, Ph.D.
GARY WEISMER, Ph.D.
*Communicative Disorders, University of Wisconsin—Madison*

JANE F. KENT
*Waisman Center, University of Wisconsin—Madison*

HOURI K. VORPERIAN
*University of Wisconsin—Madison*

JOSEPH R. DUFFY
*Department of Neurology, Mayo Clinic*

*Educational Objectives*: (1) The reader will be able to describe the major types of acoustic analysis available for the study of speech, (2) specify the components needed for a modern speech analysis laboratory, including equipment for recording and analysis, and (3) list possible measurements for various aspects of phonation, articulation and resonance, as they might be manifest in neurologically disordered speech.

**KEY WORDS:** Dysarthria; Acoustic analysis; Phonation; Articulation; Resonance; Speech production

## INTRODUCTION

Acoustic studies of dysarthria are both challenging and informative. The challenge arises because the dysarthrias can be complex disorders with potential disruptions occurring throughout the speech production system. Some disruptions may mask others, and the acoustic signal can be greatly diminished in the contrasts that are needed for precise measurements. Acoustic analysis can be informative because it affords quantitative analyses that carry potential for subsystem description and for determining the correlates of perceptual judgments of intelligibility, quality, and dysarthria type. Therefore, acoustic analysis can be

Address correspondence to Ray D. Kent, Ph.D., Waisman Center, 4th floor, University of Wisconsin–Madison, 1500 Highland Avenue, Madison, WI 53705-2280. Tel: (608) 263-7109; fax: (608) 263-0529; e-mail: <kentray@waisman.wisc.edu>.

J. COMMUN. DISORD. 32 (1999), 141–186

655 Avenue of the Americas, New York, NY 10010
0021-9924/99/$–see front matter
PII S0021-9924(99)00004-0

a valuable complement to perceptual evaluation. However, the combined use of perceptual and acoustic analyses is probably common only in specialized clinics. It is likely that joint perceptual-acoustic analysis will be increasingly common, especially with decreased costs and increased convenience of acoustic analyses. This paper considers general issues in the acoustic analysis of motor speech disorders, including types of analysis, issues regarding reliability and validity, choice of instrumentation, decisions regarding archiving, and physiological or phonetic interpretation of acoustic data. A brief review of acoustic studies of dysarthria is included, along with recommendations for future studies and clinical applications.

The dysarthrias are complex disorders of speech because they represent a variety of neurological disturbances that can potentially affect every component of speech production. In some individuals with dysarthria, the involvement may be limited to a single subsystem, for example, the larynx in a person with flaccid dysarthria affecting branches of the vagus nerve. But in many individuals with dysarthria, the disruption may be distributed over components in the respiratory, laryngeal, and supralaryngeal articulatory subsystems. Such extensive involvement is typical in diseases such as Parkinson's disease, amyotrophic lateral sclerosis, and stroke. Although perceptual judgments have been the primary means for the classification and description of the dysarthrias, questions have been raised about the reliability and validity of perceptual assessment, especially as these assessments are performed by different judges or specialists who do not have a common training in perceptual rating (Kent, 1996; Zyski & Weisiger, 1987). It has also been questioned whether perceptual judgments alone can be used to discriminate between disruptions that occur simultaneously in two or more components of speech production.

Instrumental (acoustic or physiological) assessment has often been recommended to supplement perceptual methods, in the belief that the instrumental methods will overcome some of the limitations of the more subjective perceptual assessments (Collins, 1984). In some respects, acoustic analyses can complement perceptual ratings and are particularly valuable as sources of quantitative data for clinical assessments or for tracking the effects of interventions. However, the progress in acoustic studies of dysarthria has been slow owing to several factors, including: (1) the relatively modest research effort given to neurogenic speech disorders (Strand & Yorkston, 1994), (2) the difficulty of acoustic analysis for speakers who may have phonatory disruptions, hypernasality, imprecise articulation, and other properties that confound acoustic description, and (3) the rather few published examples of broadly directed acoustic analyses of dysarthria. However, sufficient progress has been made during the last decade to synthesize and integrate acoustic analyses applied to the dysarthrias. Although a number of acoustic studies on dysarthria have been published, the great majority focus on a small set of measures and typically a very small number of subjects.

The present paper expands on earlier papers offering an overview of the acoustic analysis of dysarthria (Forrest & Weismer, 1997; Weismer, 1984). Specifically, this paper (1) summarizes recent or potential applications of acoustic analysis to dysarthric speech, and (2) proposes particular acoustic analyses that may be useful in a standardized assessment of disordered speech and voice. The issues considered are relevant to both research on dysarthria and the clinical use of acoustic analysis as a supplement to perceptual assessments. Information is included on commercially available systems designed for the acoustic analysis of speech disorders.

Although this paper focuses on dysarthria, most of the comments pertain to any speech disorder, including: the disorders associated with deafness, structural abnormalities such as craniofacial anomalies, stuttering, and developmental phonological delay. A glossary of terms used in acoustic analysis of speech is found in Appendix 1.

## GENERAL PROCEDURES IN THE ACOUSTIC ANALYSIS OF SPEECH

Acoustic analyses may be classified broadly as time-domain, frequency-domain, and time-frequency domain analyses. A typical example of a time-domain analysis is the waveform, or energy envelope of speech. The usual frequency-domain analyses include Fast Fourier Transform (FFT) spectra, Linear Predictive Coding (LPC) spectra, and the cepstrum. The standard form of time-frequency domain analysis is the spectrogram, but waterfall spectral displays also are used for this purpose. Essentially, the time-frequency analysis is a running spectrum, that is, a series of spectra obtained at selected time intervals. Fundamental frequency also can be determined by algorithms that work in either the time or frequency domain (Hess, 1992).

More recent methods of signal processing include fractals and chaos theory. Although these methods have not been extensively applied to speech in general, they have been applied to the analysis of voice and its disorders (Baken, 1990; Herzel, Berry, Titze, & Saleh, 1994; Herzel, Berry, Titze, & Steinecke, 1995; Mende, Herzel, & Wermke, 1990; Steinecke & Herzel, 1995). One form of fractal analysis is the wavelet, which finds applications in signal processing (Akay, 1998; Bruce, 1996; Wornell, 1996) and may be useful in the analysis of normal and disordered speech. A particular advantage of wavelet analysis is that it permits flexibility in frequency and time resolution (e.g., affording good frequency resolution for low frequencies and good time resolution for high frequencies).

Because acoustic analyses can take several different forms, and because these analyses can be used for various purposes, different conceptions and goals can motivate the application of acoustic analyses. Therefore, these analyses can be considered in various ways, as discussed in the following sections.

## Deterministic versus Stochastic Data

Acoustic data can be interpreted as deterministic or stochastic. Deterministic data focus on individual spectral-temporal features that relate to an aspect of speech production. For example, formant-frequency values pertaining to a particular vowel segment may be used to infer features of lingual articulation during that segment. The objective of deterministic analysis is to infer some property of speech from individual events in the acoustic record. This kind of analysis is directed toward goals such as acoustic-to-articulatory inference for individual tokens. The data frequently pertain to specific segmental or prosodic properties of speech.

Stochastic data are statistical features or indices typically collected over long time samples. For example, calculation of a long-term average spectrum (LTAS) may be used to characterize the overall energy patterns for a particular speaking task. The LTAS may not say anything about a specific event in speech, but it describes the average energy calculated for a relatively long duration of a sample. Similarly, a histogram of vocal fundamental frequency (F0) determined from a long sample of speech does not indicate how F0 varies with individual segments such as vowels and consonants, but it portrays the distribution of values over a defined sample.

Both deterministic and stochastic data have been reported for dysarthria. Deterministic data have been reported especially for acoustic segments or features related to phonetic properties. Stochastic data are particularly useful to characterize global or long-term patterns, such as the distribution of F0 or formant-frequency values over several utterances. The two forms of data can be complementary. For example, stochastic data can describe the speaker's overall use of a particular variable, and deterministic data can show how the variable is regulated for a particular speech event.

## Dimensionality of Analysis

Acoustic analysis can be accomplished with a varying number of dimensions. The minimum is a single-dimensional analysis. The maximum number of dimensions is limited by practical factors, such as time and analysis capabilities. The difference in number of dimensions can be illustrated for voice analysis, in which both single and multiple dimensions have been advocated. One example of the former is the use of the dominant rahmonic peak of the cepstrum as an acoustic correlate for perceived voice quality (Dejonckere & Wieneke, 1992, 1993, 1994). If a single acoustic index can be identified, it promises great economy of effort, given that only a single value needs to be derived or calculated. If this single value is highly correlated with a single perceptual dimension, an overall economy of description will be accomplished. An example of a multi-dimensional analysis is the Multi-Dimensional Voice Program, or MDVP (Kay Elemetrics Corporation). One advantage of a multi-dimensional approach is that it has the potential to produce profiles of impairment that are sensitive to individual variations in a given disorder.

## Phonetic and Physiological Interpretations of Acoustic Data

For the most part, acoustic data can be interpreted with respect to particular phonetic aspects of speech or to the physiological subsystem components of speech production. An example of the former is the measurement of voice onset times for voiced and voiceless stops as an index of the consonant voicing contrast in word-initial position. An example of the latter is an acoustic measure of laryngeal function, such as mean or standard deviation of F0. The basic assumption is that an appropriately selected acoustic measure can be used as an index of some phonetic or physiological aspect of speech. Some measures have the potential for a dual interpretation as phonetic contrast and physiological subsystem. As an example, formant frequencies can be used as indexes of front-back or low-high dimensions of lingual function (i.e., phonetic features related to a particular physiological component). Several such dual-interpretation measures are discussed in a later section of this paper.

Some acoustic measures may not have an immediate and direct interpretation as either conventional phonetic aspects or as physiological aspects. Spectral moments, for example, are not fully interpretable in either realm, but have value as economical descriptions of noise spectra that can be computed algorithmically with high precision (Buder, Kent, Kent, Milenkovic, & Workinger, 1996). Perhaps future research will provide interpretive guidelines for spectral moments.

## Summary: Categories of Acoustic Analysis

Most acoustic analyses, then, can be classified (1) as to whether they are deterministic or stochastic, (2) by the number of variables used in the analysis, and (3) whether the analysis focuses on phonetic aspects, physiological components, both, or neither. Examples of a single-dimensional stochastic solution are long-term statistics on F0, or a long-term average of the magnitude of the dominant rahmonic in a cepstral analysis of a sustained vowel. These data would pertain to a general description of laryngeal function and not to specific phonetic issues. Examples of a multi-dimensional stochastic solution are long-term statistics on formant frequencies or spectral moments. An example of a multi-dimensional deterministic solution is derivation of the first four spectral moments for the frication for phoneme [*s*]. Such measures are useful primarily to describe phonetic aspects of speech, but are not highly suited to physiological components because of inferential uncertainties. An example of a single-dimensional deterministic solution is the F2-F1 difference as an index of tongue fronting for vowels. These data pertain to both a phonetic description and a physiological component description.

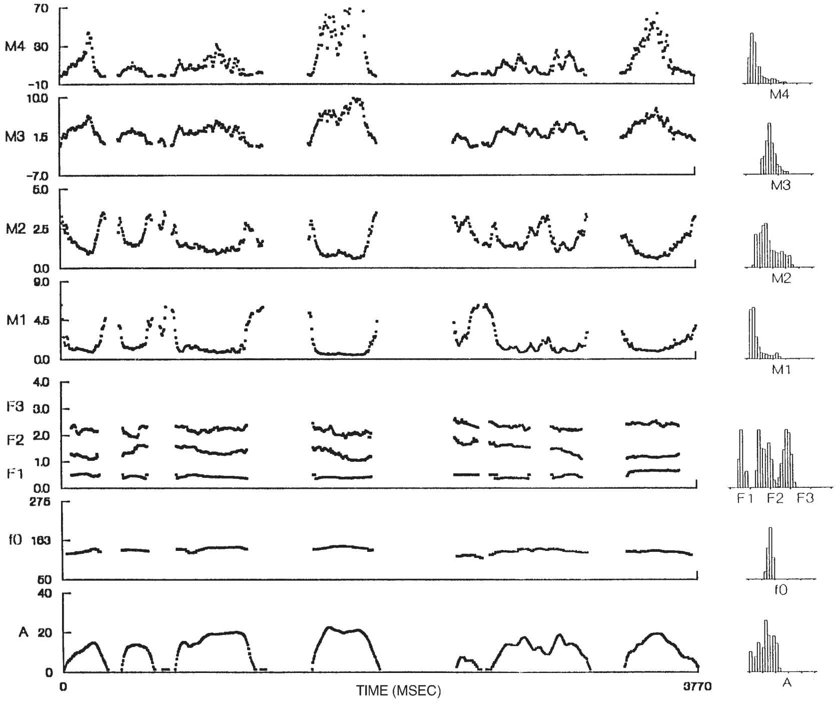


**Figure 1.** Example of a FORMOFFA plot for an individual with amyotrophic lateral sclerosis. The data are, from bottom to top, A (rms amplitude); F0 (vocal fundamental frequency); F1, F2, and F3 (the first 3 formants determined by LPC); M1 (spectral mean), M2 (spectral standard deviation), M3 (spectral skewness), and M4 (spectral kurtosis). Instantaneous values are shown in the large panel as a deterministic analysis; a stochastic summary appears at the right.

It is possible to combine deterministic and stochastic approaches. Buder et al. (1996) describes a multi-parameter acoustic analysis in which both approaches can be used. A sample display is shown in Figure 1, which includes data on rms amplitude (A), F0, frequencies of the first three formants (F1, F2, F3), and the coefficients of the first four spectral moments (M1, M2, M3, M4). Time histories for each variable are shown at the left, and cumulative histograms at the right.

## DATA QUALITY: VALIDITY, RELIABILITY AND ARCHIVAL POTENTIAL

### Validity and Reliability of Acoustic Measures

Although acoustic analysis of speech has a long history, there have been surprisingly few large-scale studies of the validity and reliability of common measurements of time, frequency, or intensity. Typically, individual research articles report only limited information on intra-judge and/or inter-judge reliability. Fortunately, because acoustic analysis has been frequently used, it is possible to establish cross-laboratory reliability for at least some measures. But even though reliability estimates have been published on time and frequency measures of normal speech, it cannot be assumed that the values from these papers can be readily generalized to dysarthric speech, or any kind of disordered speech. With high-quality recordings of normal speech, the first three formant frequencies (F1, F2, F3) can be estimated to within $+/-60$ Hz with LPC analysis (Monsen & Engebretson, 1983). With spectrography, a similar accuracy was obtained for F1 and F2, but the error for F3 increased to $+/-110$ Hz.

With respect to durational values for vowels, measurements from spectrograms and oscillograms appear to be comparable in precision, and the 95% confidence interval for careful measurements is about 10 to 25 msec (Allen, 1978). Smith, Hillenbrand, and Ingrisano (1986) concluded that temporal measures from spectrograms and oscillograms are highly similar, usually falling within 8 to 10 msec of one another. However, small consistent differences were observed for certain kinds of measures. Measurements from oscillograms tended to yield somewhat longer vowel durations and voicing during closure, while measurements from spectrograms tended to produce longer durations of consonant closure. Blomgren and Robb (1998) measured vowel steady-state durations in [Cid] syllables, using a fixed rate-of-change criterion for either the F1 or F2 frequency. They concluded from data on 40 normal speakers that durations were longer for measurements based on F1 than on F2.

Several papers have been published on factors relating to the validity and reliability of perturbation values, including the selection and placement of microphones (Titze & Winholtz, 1993; Winholtz & Titze, 1997), type of recorder (Doherty & Shipp, 1988; Perry, Ingrisano, & Blair, 1996), length of analysis window (Karnell, 1991), environmental noise (Ingrisano, Perry, & Jepson, 1998), effects of gender and trial (Jafari, Till, Truesdell, & Law-Till, 1993), and the variation among analysis systems and algorithms (Bielamowicz, Kreiman, Gerratt, Dauer, & Berke, 1996; Bough, Heuer, Sataloff, Hills, & Cater, 1996; Green, Buder, Rodda, & Moore, 1998; Karnell, Hall, & Landahl, 1995; Titze, 1994; Titze & Liang, 1993). The basic lesson to be drawn from these reports is that accurate measurements require careful attention to all phases of research—from the microphone to the recording medium and the analysis software. For the present, it seems prudent to use the studies of perturbation measures as a general guideline for other acoustic measures, which generally have not been evaluated as extensively.

Because most evaluations of measurement accuracy have pertained only to high-quality recordings of normal speech, it cannot be assumed that similar results will apply to disordered speech. Generally, studies of dysarthric speech report poorer accuracy of measurement than that indicated for normal speech.

Of course, this result is not surprising, given the reduced contrast typical for disordered speech, which may be dysphonic, nasalized, and otherwise different from normal speech. The issue of precision is particularly important to the observation that at least some types of dysarthric speech are more variable than neurologically normal speech. Care must be taken to distinguish variability intrinsic to the measurement process from variability that reflects the actual speech behaviors under examination.

## Archival Considerations

Until fairly recently, most recordings of clinical speech samples were made with analog tapes, using either reel-to-reel or (more commonly) cassette recorders. But the availability of economical digital audio tape (DAT) recorders is eclipsing the use of analog recorders. Typically, DAT recorders afford sampling rates of 32, 44.1, or 48 kHz, with the rate of 44.1 kHz being most frequently used in speech research. This sampling rate allows the recording of signal frequencies up to nearly 20 kHz, which is very generous compared to the historic 8-kHz limit of conventional spectrography. This expanded frequency range, coupled with quantization at 16 or 32 bits, gives DAT recorders a very high fidelity of recording. In addition, DAT recorders are easy to use, having user operations similar to those of analog recorders. Consequently, DAT recorders can replace analog machines with relative convenience and economy.

However, it should be noted that because DAT is a metal-particle tape, it is subject to eventual deterioration. In one study of the corrosion of metal-particle and metal-evaporated tapes, it was noted that the stability and reliability became questionable when temperature and humidity were not carefully controlled (Speliotis & Peter, 1991). For magnetic media in general, deterioration can be detected within 5 to 8 years after recordings are made (Leek, 1995). Despite the impression held by some that DAT affords a relatively permanent storage of recorded information, it should not be assumed that these tapes will guarantee high-fidelity archiving of speech samples. Although control of temperature and humidity will extend the accuracy of the data, errors will ultimately occur. But for short-term applications, DAT is certainly a major advance over analog tape. For long-term storage without loss of data, the best alternatives to consider are the recordable compact disc (CD-R), the compact disc rewritable (CD-RW), or the MiniDisc.

Very few archival recordings of dysarthric speech are generally available. The Whitaker database (Deller, Lilu, Ferrier, & Robichaud, 1993) consists of 81 words recorded from 6 adult males with cerebral palsy (two spastic, two athetoid, one spastic athetoid, and one spastic ataxic). A single normal speaker is included as a control. The speech was high-pass filtered at 75 Hz, low-pass filtered at 4.7 kHz, and sampled at 10 kHz with 12-bit conversion. The database can be obtained through the electronic mail network. This database is useful for testing speech recognition systems applied to disordered speech, but it is limited to adult male speakers and the neuropathology of cerebral palsy. The bandwidth restriction limits the opportunities to examine high-frequency elements such as noise bursts or frication segments. Aronson's (1993) tape recording of several types of dysarthria was intended primarily as an aid to instruction in the classification of dysarthrias according to the Darley et al. (1969a, 1969b) system. The speech samples include vowel prolongation, syllable repetition (diadochokinesis), and passage reading. The speakers were selected both to represent various types of dysarthria and to be reasonably controlled with respect to severity. The samples are available on analog cassette tape.

## BASIC REQUIREMENTS FOR AN ACOUSTIC SPEECH LABORATORY

A basic but powerful laboratory for effective recording and analysis of dysarthric speech includes the following components:

1. A miniature head-mounted condenser microphone ensures high-quality recordings even when the subject's head moves, as might occur particularly in patients with head tremor or dystonia. Winholtz & Titze (1997) report on such a microphone that appears to be highly suitable for clinical and research recordings.
2. A DAT recorder with a 44.1-kHz sampling rate and either 16 or 32-bit conversion is an excellent choice for general applications (although the suitability of DAT for archival purposes is questionable, as noted earlier). A 16-bit conversion permits 65,536 levels of amplitude to be represented in the digitized speech sample.
3. Several multi-purpose speech analysis systems that afford time-domain and frequency-domain analyses are available (Read, Buder, & Kent, 1992; Gopal, 1995). Prices vary considerably, but some systems can be purchased for well under $1,000 while others will cost more than $5,000. Because these systems can be complementary in their features, it is a good idea to consider the purchase of two or more, depending on laboratory needs and the available computer platform.
4. High-quality headphones or speakers are a definite asset. Gopal (1995) lists recommendations for headphones in a speech research laboratory.
5. A large screen monitor is not a necessity, but can be very helpful in visualizing multiple displays on one screen and in detecting subtle acoustic events. The advantage of a large screen is especially appreciated when the user must perform a large number of analyses.

6. A quiet environment for obtaining acoustic recordings is critical. This can be difficult in settings affected by voice paging systems, heating and air conditioning equipment, and other environmental sources of noise. Ideally, a sound-treated recording booth should be used. A local source of noise that is frequently overlooked is fan noise from a personal computer, which can affect acoustic analyses (Ingrisano et al., 1998).

Information sources on basic issues include: digitization (Kent & Read, 1992; Venkatagiri, 1996); computer-assisted analysis (Gopal, 1995; Kent & Read, 1992), synthesized speech (Venkatagiri, 1996); and automatic speech recognition (Kent & Read, 1992).

## ASSOCIATIONS OF ACOUSTIC MEASURES WITH ASPECTS OF SPEECH MOTOR CONTROL

The following sections discuss acoustic measures related to major aspects of the motor control of speech. Typical (or potential) measures are listed under each category, and published applications to dysarthria are reviewed briefly.

### Phonation or Voice Quality (Laryngeal Subsystem)

**Suggested measures:** F0 statistics (mean, mode, range, standard deviation, interquartile ranges, etc); F0 contour for individual utterances; perturbation measures (jitter and/or shimmer); harmonic to noise ratios; spectral tilt and spectral energy ratios; wow, tremor, and flutter. (See Buder, in press).

**Voice quality in dysarthria: General comparisons.** Because a disorder of voice quality is common in dysarthria (Darley et al., 1969a, 1969b; Duffy, 1995), it is important to identify the acoustic correlates of these disorders in dysarthric subjects. However, the success of this effort has been marginal. In a comparison of subjects with Parkinson's disease and normal controls, Ludlow, Coulter, and Gentges (1983) did not observe a statistically significant difference in jitter. Zwirner, Murry, and Woodson (1991) reported that acoustic measures did not distinguish among three types of neurological disease (Parkinson's disease, Huntington's disease, and cerebellar ataxia) and, for that matter, did not distinguish these clinical subgroups from normal controls. Similarly, Kent, Kim, Weismer, Kent, Rosenbek, Brooks, and Workinger (1994) concluded that their acoustic measures (F0, jitter, shimmer, and signal-to-noise ratio) did not distinguish among patients with three neurological diseases (amyotrophic lateral sclerosis, Parkinson's disease, and cerebrovascular accident) and did not separate the clinical subgroups from normal controls. It appears that long-term measures of phonatory instability (such as the standard deviation of F0) may hold more potential than perturbation measures such as jitter and shimmer (Zwirner et al., 1991). Similar long-term measures have been used to demonstrate significant linear declines in longitudinal studies of subjects with Parkinson's disease (King, Ramig, Lemke, & Horii, 1994).

A complicating factor in the study of voice disorder is the heterogeneity of the impairment in subjects with the same neurological classification. Strand, Buder, Yorkston, and Ramig (1994) reported differential phonatory characteristics in four women with amyotrophic lateral sclerosis (ALS) who had initial bulbar signs and progressive deterioration of phonation. Strand et al. questioned whether group data on dysarthria in ALS or other neurogenic speech disorders may simply mask the large variability among patients. Further evidence of this variability was reported for stroke by Murdoch, Thompson, and Stokes (1994). Only about half of their subjects with upper motor neuron disease demonstrated hyperfunctional features such as elevated subglottal air pressure, increased glottal resistance, and decreased laryngeal airflow. The other half of their subjects presented with features of hypofunctional laryngeal activity.

The influence of speaker sex, age, and race also must be considered in studies of voice disorder in dysarthria. Differences in phonatory function between men and women have been reported for amyotrophic lateral sclerosis (Kent et al., 1994) and Parkinson's disease (Hertrich & Ackermann, 1995). Laryngeal function—and supralaryngeal function as well—can be affected by aging (Linville, 1996; Weismer & Liss, 1991). Because dysarthria often occurs in neurological diseases that are more common in the elderly, the most appropriate normal control subjects in dysarthria research are older adults. Similarly, clinical evaluation should use age-appropriate (and sex-appropriate) normative standards. Although it is difficult to make confident conclusions regarding the effect of race on voice, some studies indicate that race is a relevant factor in establishing norms for aspects of vocal function (Hudson & Holbrook, 1981, 1982; Ryalls, Zipprer, & Baldauff, 1997; Walton & Orlikoff, 1994).

**Harmonic versus noise energy.** Several studies of voice disorders have drawn attention to measures of the relative amounts of harmonic and noise energy in vowel phonation. Because these measures have often correlated well with perceptual ratings of voice quality, and because they can be computed fairly easily, they are good candidates for the study of voice disorders in individuals with dysarthria. Only a few studies are mentioned here, but the relevant literature is substantial. Hiraoka, Kitazoe, and Ueta (1984), in an harmonic-intensity analysis of normal and hoarse voices, concluded that hoarse voices have a prominent fundamental frequency intensity compared with harmonics in the voice spectrum. They defined a measure of relative harmonic intensity, $H_r$, obtained from a stable portion of vowel /a/, as the intensity of the 2nd and higher harmonics expressed as a percentage of the total voice intensity. Most normal voices were found to have an $H_r$ larger than 67.2%, whereas most hoarse voices had values below this criterion. Dejonckere and Wieneke (1992, 1993, 1994) reported on the clinical value of the magnitude of the main cepstrum peak. They noted that in pathological voices, the magnitude of the main cepstrum peak was negatively correlated with phonation flow, the relative noise above 6 kHz, the jitter ratio, and the perceptual evaluation of the grade of hoarseness (Dejonckere & Wieneke, 1992, 1993). It was also reported that the magnitude of the main cepstrum peak was better than degree of aperiodicity or excess of high-frequency noise for demonstrating functional improvement following surgery (Dejonckere & Wieneke, 1994).

**Wow, tremor, and flutter.** Tremor (defined as an oscillatory motion of part of the body) can be observed in several neurological disorders and can be manifest in speech, especially during the task of sustained phonation (vocal tremor). Some writers distinguish wow (oscillation of 1–2 Hz), tremor (oscillation of 2–10 Hz) and flutter (oscillation of 10–20 Hz) (Hartelius, Buder, & Strand, 1997). Acoustic methods permit a quantification of these oscillatory phenomena that is not easily achieved by perceptual methods. Generally, the frequency of oscillation varies with the type of disorder, as summarized in Figure 2. The tremor in both cerebellar disease and Parkinson's disease is relatively slow, in the range of about 3 to 7 Hz (Ackermann & Ziegler, 1991a; Ludlow, Bassich, Connor, & Coulter, 1984; Philippbar, Robin, & Luschei, 1989; Ramig, Scherer, Titze, & Ringel, 1988; Ramig & Shipp, 1987). These tremor frequencies are low compared to both normal tremor and the flutter that occurs in some patients with amyotrophic lateral sclerosis (Aronson, Ramig, Winholtz, & Silber, 1995). Low-frequency wow has been reported in particular for individuals with multiple sclerosis, typically combined with higher-frequency oscillations (Hartelius et al., 1997). Oscillatory behaviors in neurological disorders can be complex, and although one frequency of oscillation may dominate, other co-occurring frequencies may exist (Boutsen, Duffy, & Aronson, 1998; Hartelius et al., 1997).

In one of the earliest quantitative analyses of vocal tremor, Brown and Simonson (1963) reported data for 24 patients. Tremor during sustained vowels was analyzed with respect to both the frequency of tremor and the percentage change in amplitude, as determined from oscillographic records. The frequency of essential vocal tremor varied from 4 to 9 Hz, with the most common value falling between 5 to 7 Hz. The percent change in amplitude varied from 40 to 100%, with most patients having values between 60 and 80%. The same measurement principles are useful today, but analysis methods have been refined to include multi-parameter analyses (Ackermann & Ziegler, 1991a), FFT spectra (Hartelius et al., 1997) and tremor cancellation (Gath & Yair, 1988; Winholtz & Ramig, 1992).

**Multi-dimensional voice analyses.** A broad analysis of voice is possible with nearly automatic procedures, including the Multi-Dimensional Voice Profile™ (MDVP). An example of a MDVP analysis of the voice of an individual with dysarthria is shown in Figure 3. The lightly shadowed circle partly visible in the illustration represents normative values for each parameter (it

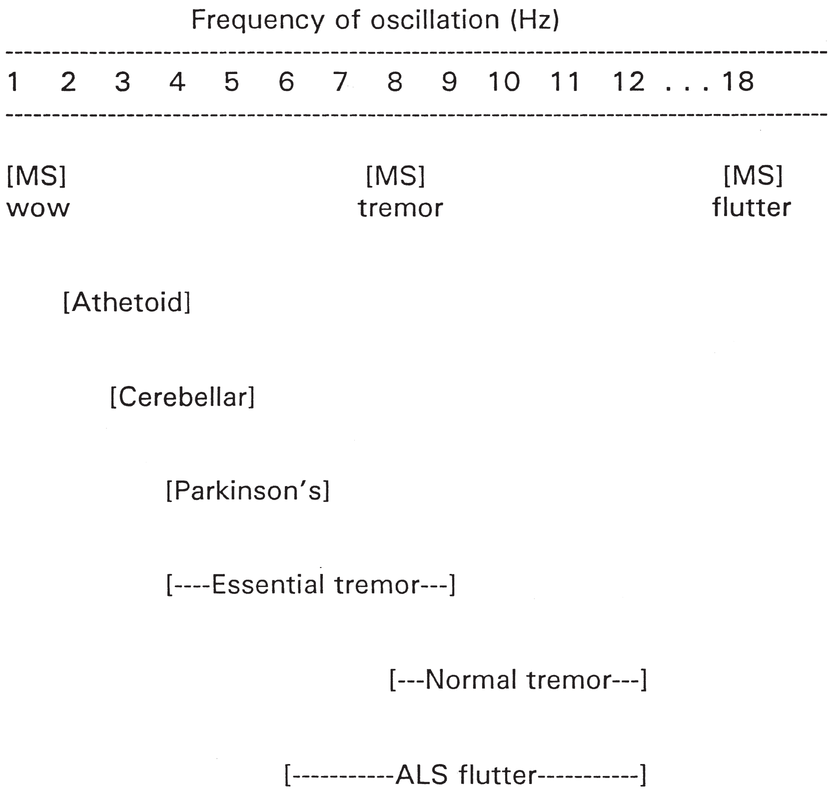


**Figure 2.** Oscillation frequencies associated with various neurological disorders.

should be noted that these normative values may not be appropriate for subjects of different ages). For this patient, abnormal values occur for several parameters, including absolute jitter (Jita), percentage jitter (Jitt), relative average perturbation (RAP), pitch perturbation quotient (PPQ), smoothed pitch perturbation quotient (sPPQ), fundamental frequency variation (vF0), shimmer in dB (ShdB), shimmer percent (Shim), amplitude perturbation quotient (APQ), smoothed amplitude perturbation quotient (sAPQ), peak-amplitude variation (vAm), soft phonation index (SPI), and amplitude tremor intensity index (ATRI).

## Vocal Tract Function for Vowel Articulation: Spectral Correlates

**Suggested measures:** absolute values of F1, F2, and F3 frequencies; F1–F0 difference value; F2–F1 difference value; F2/F1 planar area for vowel triangle or quadrilateral; formant frequency fluctuation.

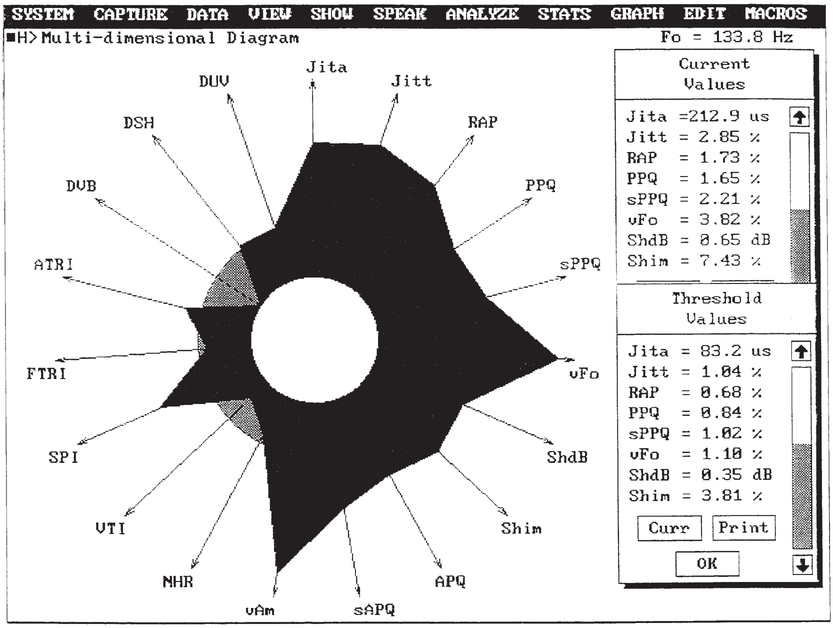


**Figure 3.** Example of MDVP diagram for voice analysis. See text.

**Acoustic-articulatory relationships.** Vowel articulation is usually indexed acoustically by measures of formant frequencies, customarily the first two or three (F1, F2, F3). Because formant frequencies vary with the length of the speaker's vocal tract, and therefore with speaker age and sex, data on absolute frequencies must be used with consideration of speaker characteristics. Typically, vowel formant-frequency data have been reported as absolute values for a specific age-sex group, but some reports have used vowel plane area (Turner, Tjaden, & Weismer, 1995) or isovowel lines (Kent, 1979; Kent, Netsell, & Abbs, 1979).

A general rule in acoustic-articulatory relationships is that F1 frequency varies inversely with tongue height and F2 frequency varies inversely with tongue advancement. An alternative interpretation is that the F1–F0 difference varies inversely with tongue height and the F2–F1 difference increases with tongue advancement. These difference values can be expressed in absolute frequency or as some transformation, such as logarithms or Bark (Syrdal & Gopal, 1986). One caveat in the use of the F1–F0 difference in dysarthric speech is that F0 is subject to many influences and may be erratic in some samples of dysarthria. Particularly when laryngeal function is highly variable, or when the subject is dysphonic, the F1–F0 difference is of questionable value. Articulatory-acoustic relations for vowels can be summarized as follows:

*Advanced or front*—high F2, large F2–F1 separation
*Retracted or back*—low F2, small F2–F1 separation
*Low*—high F1 or large F1–F0 separation
*High*—low F1 or small F1–F0 separation
*Centralized*—all formant values converge on targets for schwa
*Reduced vowel contrasts*—reduction of planar for vowel triangle or quadrilateral
*Lip rounding*—all formants decrease in frequency

**Vowel Production in Dysarthria.** The most frequently reported abnormalities of vowel production in dysarthria include: centralization of formant frequencies (Ziegler & von Cramon, 1983a, 1983b, 1986b), reduction of the acoustic plane area for vowels (Turner et al., 1995), and abnormal formant frequencies for high vowels and front vowels (Watanabe, Arasaki, Nagata, & Shouji, 1994). Figure 4 shows how Turner et al. determined the area of the F1–F2 quadrilateral by summing the areas of the triangles T1 and T2.

Instability in vowel formant pattern can also be of interest. Formant-frequency fluctuation is a variability in formant pattern in sustained vowel phonation or other vowel steady states (Gerratt, 1983; Robb, Blomgren, & Chen, 1998). This fluctuation can be used to document involuntary vocal tract movements in disorders such as tremor, chorea, or dystonia. Computation of this

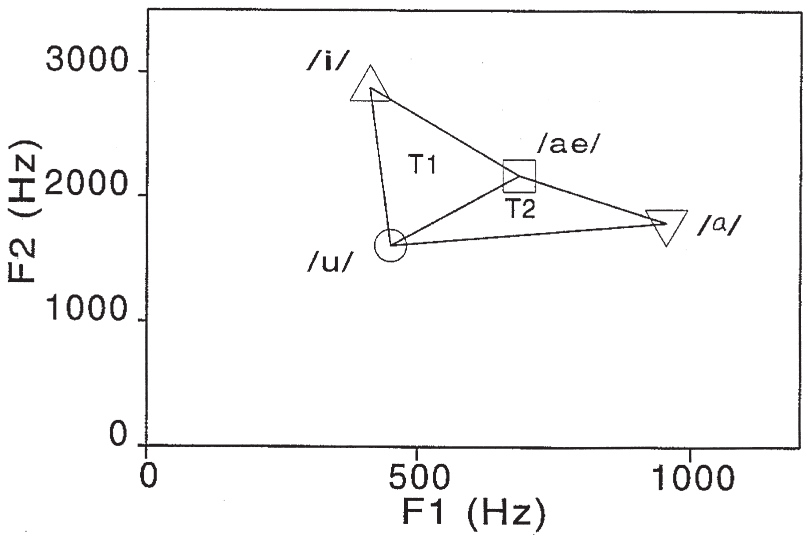


**Figure 4.** Example of area determination for vowel quadrilateral by summing areas for constituent triangles T1 and T2 (from Turner et al).

measure can be simplified with LPC formant tracking. Figure 5 shows a wide-band spectrogram and waveform for an attempted vowel phonation by a subject with severe vocal and oromandibular tremor. The tremor is manifest as conspicuous modulations of the waveform amplitude and as variations in the vowel formant pattern (including the higher formants F4, F5, and F6). In addition, the subject's phonation is occasionally interrupted by glottalized intervals (indicated by the arrows).

## Vocal Tract Function for Consonant Articulation: Spectral Correlates

**Suggested measures:** spectrum of stop burst or fricative noise, spectral moments computed for stop burst or fricative noise, formant transitions for CV and VC transitions.

**Acoustic-Articulatory Relationships.** Because consonants are a complex class of sounds, there is no single measure that can be applied across the class. A useful distinction can be made between sonorants and nonsonorants. The sonorants (glides, liquids, and nasals) can be described by patterns of formants

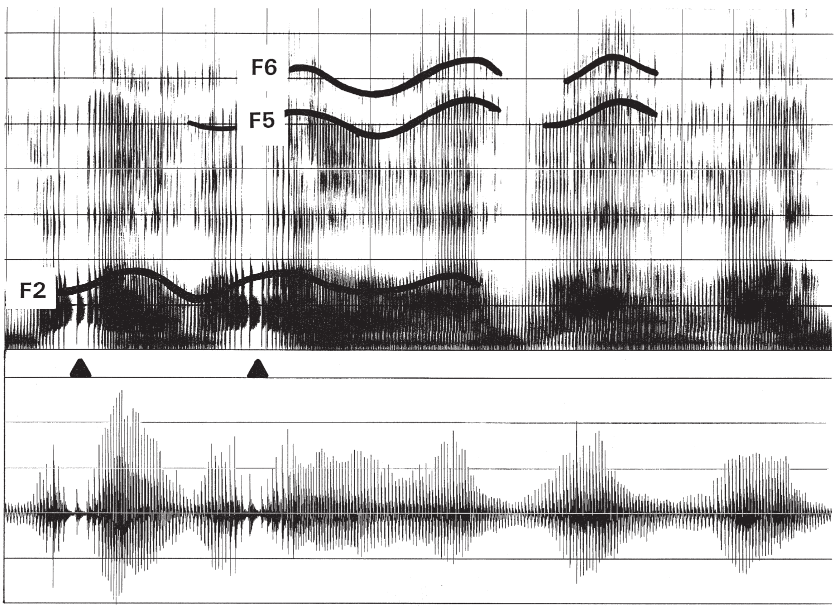


**Figure 5.** Wide-band spectrogram and waveform for phonation of /a/ by a woman with severe vocal and oromandibular tremor. Arrowheads point to intervals of glottalization, and the thick lines are estimates of the F2, F5, and F6 frequencies. See text for discussion.

and antiformants in either steady-state or transitional segments. Therefore, data are similar to those used for vowels. The nonsonorants (stops, affricates, and fricatives) involve some kind of frication event: a burst or transient noise for stops, a brief noise interval for affricates, and a longer noise interval for fricatives. Published information on dysarthria emphasizes nonsonorants, and the following discussion reflects this emphasis.

**Burst and frication noise in dysarthria.** A natural step in the study of nonsonorant articulation is characterization of the noise. Durational measures are typically straightforward, but spectral measures are not. Spectral properties of stop and fricative noise have been studied in both dysarthria and aphasia (Forrest & Weismer, 1997; Harmes, Daniloff, Hoffman, Lewis, Kramer, & Absher, 1984; Shinn & Blumstein, 1983; Tjaden & Turner, 1997; Ziegler & von Cramon, 1986b). There is no generally accepted means of summarizing noise spectra as a small number of quantitative indexes. Spectral moments may suffice for many purposes, but they have not been used extensively in the study of dysarthria. The first moment (mean) gives the center of gravity for the noise and appears to be sensitive to certain kinds of fricative misarticulation. The most likely deviation in dysarthric samples is a reduction of the first moment (Ziegler & von Cramon, 1986b). The second moment (standard deviation) expresses the distribution of energy around the mean. The third moment (skewness) describes the symmetry of the distribution, and the fourth moment (kurtosis) pertains especially to the peakedness of the energy distribution (assuming a normal distribution). In one of the very few studies in which spectral moments were applied to dysarthria, Tjaden and Turner (1997) compared fricative spectra in the speech of persons with ALS and the speech of neurologically normal subjects. The difference in frequency for the first moment of [*s*] and [ʃ] was correlated with perceptual judgments of consonant precision. The higher moments were associated with more complex patterns. It appears that the first moment (spectral mean) is a useful index of fricative production, particularly when combined with measures of the duration and energy of the noise segment.

**F2 slope index in dysarthria.** An acoustic index that relates to articulatory dynamics is the slope of the F2 transition in CV sequences (Weismer & Martin, 1992; Weismer, Martin, Kent, & Kent, 1992). It was shown that in men with ALS, the averaged F2 slope of a small group of test words (called the F2 slope index) was correlated with overall speech intelligibility scores (Kent, Kent, Weismer, Sufit, Rosenbek, Martin, & Brooks, 1990). In addition, it was reported that the longitudinal decline in overall speech intelligibility in a woman with ALS paralleled a consistent decline in F2 slopes (Kent, Sufit, Rosenbek, Kent, Weismer, Martin, & Brooks, 1991).

**Energy measures in dysarthria.** The precision of stop consonant production can be determined in part by measures of the acoustic energy during the intended occlusive phase, or stop gap (Ackermann & Ziegler, 1991b; Weismer, 1984). In general, normal production of a voiceless stop consonant is associated with a virtually silent stop gap. But some dysarthric speakers, especially those with Parkinson's disease, tend to produce energy during the gap. This energy is typically one of two forms: turbulence noise (spirantization) generated at the site of oral constriction because of an incomplete occlusion, and voicing energy, which often occurs because of poor coordination between laryngeal and supralaryngeal actions. The spectrum of the energy is usually quite distinctive between these two types of error. The stop-gap energy is especially likely for syllables that do not receive primary stress (Ackermann & Ziegler, 1991b), but—especially in subjects with severe impairment of articulation—the energy may be noted frequently for consonants in syllables of varying stress levels and even during a stress-uniform task such as syllable repetition (diadochokinesis).

## Velopharyngeal Function

**Suggested measures:** formant frequency shifts, reduced formant amplitudes, increased formant bandwidths, presence of nasal formants and antiformants.

**Acoustic-articulatory relationships.** Acoustic correlates of hypernasal speech are not easily summarized, as several different factors have been described whose importance may vary with speaker and phonetic context. Nasalization can be identified acoustically from some combination of the following: presence of nasal formant, reduction of overall energy, increased bandwidth of formants, shifts in formant frequencies, and presence of antiformants (Kent & Read, 1992). Therefore, one approach to the acoustic study of nasalization is to identify two or more expected correlates of nasalization (Ansel & Kent, 1992). Recent studies of nasalization in both normal and disordered speech point to possible improvements in the detection and measurement of nasality (Chen, 1997; Plante, Berger-Vachon, & Kauffman, 1993). Chen studied the use of two amplitude difference values based on the following measures: P0, the amplitude of an extra peak in the low frequencies; P1, the amplitude of an extra peak located between the first 2 formants; and A1, the amplitude of the first formant. The difference values were A1-P1 and A1-P0. The A1-P1 difference averaged more than 10 dB for English speakers' productions of oral versus nasalized vowels. Plante et al. (1993) demonstrated that particular LPC coefficients were sensitive to the presence of nasalization in vowels produced by children. If the same sensitivity could be demonstrated in adults, then LPC spectra may be a relatively simple means for the study of velopharyngeal function in dysarthria.

## Coordination of Laryngeal and Supralaryngeal Activity

**Suggested measures:** voice onset time, amplitude of H1 during fricative segment.

**Acoustic-articulatory relationships.** The basic assumption in these measures is that two or more acoustic events are consistently associated with an underlying physiological pattern. For example, in the case of voice onset time (VOT) for syllable-initial stops, it is assumed that the acoustic interval between the burst and the onset of periodic energy corresponds to the physiological interval between release of the consonantal constriction and the onset of vocal fold vibration.

**Voice onset time in dysarthria.** VOT is perhaps the most frequently used index of subsystem coordination, and a relatively large amount of data has been published on VOT in normal speech and several varieties of disordered speech. Despite this comparatively large database, several details remain to be addressed, including the possibility of differences related to sex, age, race, and phonetic context (Nieman, Klich, & Shuey, 1983; Ryalls et al., 1997; Swartz, 1992; Weismer, 1979). Questions have been raised about the degree to which VOT in itself is a satisfactory index of the coordination of laryngeal and supralaryngeal events. Caruso and Burton (1987) commented that VOT differences have not been found in several studies that compare normal speech with speech disorders thought to be associated with laryngeal-supralaryngeal discoordination (Kent et al., 1979; Metz, Conture, & Caruso, 1979; Watson & Alfonso, 1982). It is probably wise to supplement VOT data with other acoustic measures pertaining to the stop gap, voiceless interval, and aspiration (Klatt, 1975; Weismer & Fromm, 1983). A composite measure, the laryngeal devoicing gesture (LDG), is defined for voiceless stops as the sum of closure duration and the VOT (Lisker & Abramson, 1964).

VOT data have been reported for several neurogenic communication disorders including aphasia (Blumstein, Cooper, Goodglass, Statlender, & Gottlieb, 1980; Blumstein, Cooper, Zurif, & Caramazza, 1977), apraxia of speech (Freeman, Sands, & Harris, 1978; Itoh, Sasanuma, Tatsumi, Murakami, Fukusako, & Suzuki, 1982; Ziegler, 1987), and dysarthria (Caruso & Burton, 1987; Farmer, 1980). Figure 6 shows three examples of VOT distributions for a voiced-voiceless cognate pair. Part A shows the expected result for citation-style speech produced by a normal talker. The VOT values for the voiced and voiceless items form nonoverlapping distributions with a boundary at about 25 ms. Part B illustrates a result in which both the voiced and voiceless targets are associated with large ranges of VOT values and the two distributions overlap somewhat. Part C depicts a situation in which the two distributions overlap considerably and together occupy a restricted range of VOT values.

**H1-frication index in dysarthria.** Another measure pertinent to coordination is the amplitude of H1 (first harmonic) during an intended voiceless fricative segment. A coordinative difficulty is indicated when H1 invades the fricative interval. This measure apparently has not been reported for dysarthric speech, but procedures described by Pirello, Blumstein, & Kurowski (1997) for normal speech could be used with disordered speech as well.

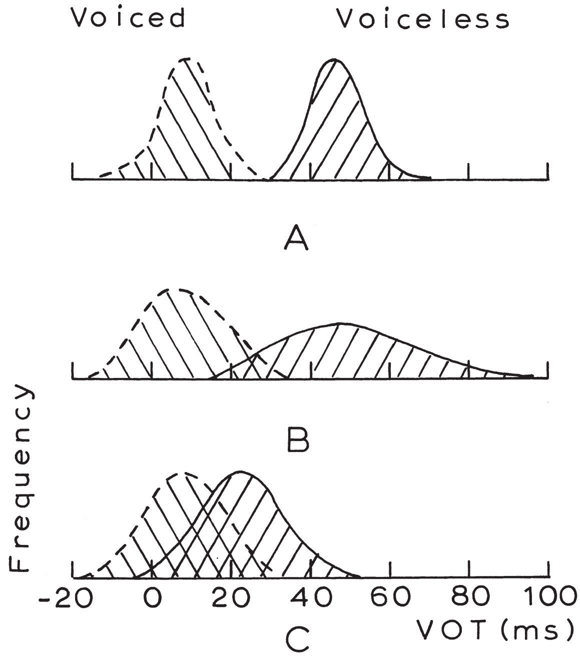


**Figure 6.** Examples of VOT distributions for voiced and voiceless stops.

## Syllable Timing

**Suggested measures:** durations of syllables and intersyllable pauses.

**Diadochokinesis: Acoustic-articulatory relationships.** Even simple forms of acoustic analysis can be used to examine the temporal structure in syllable sequences. The relevant information can be derived from measurements of the waveform or the energy envelope of speech.

**Syllable timing in dysarthria.** At least for some aspects of syllable timing, acoustic methods may provide information not easily obtained from auditory evaluation. For example, in their original description of spastic dysarthria, Darley et al. (1969a, 1969b) concluded that syllables were repeated at a slower than normal rate but with normal rhythm. However, in an acoustic study Portnoy and Aronson (1982) observed an abnormal variability of syllable rhythm in a sample of 30 subjects with spastic dysarthria. They suggested that "current assumptions about the dysarthrias that have been determined from perceptual judgments be reexamined from the perspective of quantitative analyses" (Portnoy and Aronson, p. 327). It has also been shown that quantitative analysis of diadochokinesis reveals profiles of performance that are to some extent specific to the type of neurological disorder. Ackermann, Hertrich, and Hehr

(1995) concluded from a multivariate analysis that subjects with Parkinson's disease generally produced lengthy repetition trains characterized by incomplete closures whereas subjects with Friedreich's ataxia produced trains with only a few syllables at a slow rate and with complete closures.

Syllable timing has been a particular focus in the study of ataxic dysarthria which is said to have a scanning or staccato rhythm. Abnormalities of rhythm also are frequently noted for the syllable repetition task. The usual conclusion is that the rate of repetition is slow and the interval between syllables is irregular (Boutsen, Bakker, & Duffy, 1997; Cisneros & Braun, 1995; Gentil, 1990; Kent et al., 1997; Kojima, Shimoyama, Ninchoij, & Uemura, 1989; Portnoy & Aronson, 1982; Ziegler & Wessel, 1996). Duffy (1995) has recommended the syllable repetition task as being particularly sensitive to cerebellar dysfunction. Confirming evidence was reported by Ziegler and Wessel (1996), who concluded that maximum syllable repetition rate predicted with reasonable success the perceived severity of ataxic dysarthria (accounting for nearly 70% of the variance in severity ratings and nearly 60% of the variance in intelligibility ratings).

**Scanning index (SI) in dysarthria.** Ackermann and Hertrich (1994) proposed an index of temporal structure that can be used to evaluate the presence of a scanning pattern of speech. The proposed index is defined as:

$$SI = (S_1 \times S_2 \times \ldots \times S_n) / [(S_1 + S_2 + \ldots + S_n)/n]^n$$

Ackerman and Hertrich explain SI as follows: "Provided that all of the [*n*] syllables have equal lengths, the index amounts to unity. In any other case, especially if one syllable is considerably shorter than the other ones, this measure will be $<1$." (Ackermann and Hertrich, 1994, p. 80). The general prediction that speakers with ataxic dysarthria would show a tendency toward isochronicity, or scanning speech, was not confirmed by either Ackermann and Hertrich or by Kent et al. (1997). Apparently, the dysarthria in cerebellar disease is not associated with a uniformly isochronic syllable pattern. The appearance of even one markedly shortened syllable can effect the value of SI. The perceived scanning pattern may be associated more with unusual patterns of vowel duration, such as the infrequent appearance of unstressed vowels (Kent et al., 1979). It may be most useful simply to report the distribution of vowel or syllable durations.

## Sound Segment Timing

**Suggested measures:** segment durations.

**Acoustic-articulatory relationships.** The essential relationship is simple, at least at first impression. But segment durations measured acoustically need consistent guidelines for physiological interpretation. For example, Blomgren and Robb (1998) discussed issues in the measurement of vowel steady states, noting that the values differed according to which formant was used as a criterion measure. Probably the most serious problem is that segments that are typically well defined in normal speech can be obscured in dysarthric speech.

**Segment durations in dysarthria.** A number of studies have examined segment durations in dysarthric speech (Caruso & Burton, 1987; Hertrich & Ackermann, 1997; Kent et al., 1979; Weismer, 1997). Because most individuals with dysarthria have slow speaking rates, it is expected that segment durations will be longer in dysarthric samples than in normal control samples. However, the issue becomes more interesting because the lengthening affects some segments more than others. An understanding of these differences is relevant to the explanation of reduced intelligibility and perhaps to hypotheses about underlying disorders of neural regulation. It has been concluded from some studies of dysarthric speech that vowel and consonant segments were lengthened relative to their durations in normal speech, but VOT values were not different from those for normal controls (Caruso & Burton, 1987; Kent et al., 1979).

## Qualitative Analysis

The great majority of studies reviewed here have focused on a small set of measures (e.g., VOT, formant frequencies, segment durations) that were summarized and evaluated using standard quantitative techniques. These quantitative summaries are in the form of means, standard deviations, ranges, and so forth, and the evaluation of the summary data usually takes the form of statistical tests. Most of what is known about the acoustic aspects of motor speech disorders has been learned from these kinds of analyses, but there are also non-quantitative approaches to understanding dysarthric speech. These *qualitative* analyses usually require the careful examination of acoustic records of many individual utterances, which may lead to novel discoveries about the speech production deficit in dysarthria. This careful examination may occur during the course of more standard measurement of a specific acoustic variable (such as a formant frequency), or may be undertaken on its own as a worthy scientific endeavor. Qualitative analyses might be driven by perceptual hypotheses or observations. One of the dangers of automated, computer-based analysis of dysarthric speech is that it may discourage the careful examination of qualitative features, and therefore cause the user to overlook phenomena with clinical and theoretical import. A phenomenon identified by means of qualitative analysis may even lead to the formulation of a hypothesis that can be tested using the more standard quantitative techniques. What makes this sequence of scientific activity (qualitative observation→hypothesis→quantitative test) more than a truism is that the qualitative analysis may direct the use of acoustic measures that are in the “standard” arsenal and may never have been considered in the absence of the qualitative analysis. A few examples are given to illustrate these points.

As reviewed throughout this paper, dysarthric speakers frequently have more variable acoustic measures than persons with normal neurological status. This greater variability has often been regarded as the logical consequence of a neurologically-impaired system, as if the impairment increases the “noise” in speech production performance. In some disorders—such as athetosis—the speech production deficit, and consequently the speech acoustic output, has been considered to be the direct result of random fluctuations in muscle tone (Darley et al., 1975). Liss and Weismer (1992) examined formant trajectory plots of multiple repetitions of several words produced by individuals diagnosed with apraxia of speech and ataxic dysarthria, and compared these to corresponding plots obtained from the speech of neurologically-normal speakers. When the trajectories from the multiple repetitions of neurologically-normal speakers were superimposed, the traces appeared to fit on top of one another, as if the trajectory for any one repetition could be interchanged for that of any other repetition. An example for a normal subject is shown in the top part of Figure 7. The consistency of every aspect of the trajectories (starting frequency, slope of major transition, range of frequencies covered over the major transition, and so forth) was remarkable evidence that normal speakers have fine precision in reproducing frequently-used vocalic nuclei.

The superimposed plots of multiple formant trajectories from the patients, on the other hand, appeared to be disorganized in all aspects, as if the speech production characteristics of any one repetition were entirely different from those of any other. An example from an individual with apraxia of speech is shown in the lower part of Figure 7; note the variability in the different trajectories labeled with trial number. Plots like this could easily have been interpreted as a prime example of a speech production mechanism made extremely noisy by neurological disease. However, a qualitative analysis showed that much of the apparent disorganization of these repeated trajectories could be explained by fluctuation in the extent to which the vocalic nucleus and the preceding consonant were coarticulated. In other words, the variation seen in the formant trajectory plots appeared to have a systematic basis in a timing control variable, which is quite different from a view that emphasizes “random control” (noise). The observation of systematicity observed in Liss and Weismer (1992) was developed into a quantitative hypothesis, described in detail in Weismer, Tjaden, and Kent (1995).

Two other qualitative phenomena that have not been adequately studied merit brief discussion: (1) As reviewed earlier, acoustic characteristics of phonation have been described using jitter, shimmer, and signal-to-noise ratio, as well as other well-known voice measures. Qualitative analysis, using spectrographic displays, reveals another important aspect of voice production in many individuals with dysarthria—the tendency for sudden, short-lived shifts in phonatory behavior for vowels. These phonatory changes, typically displayed spectrographically as instantaneous (usually downward) shifts in fun-

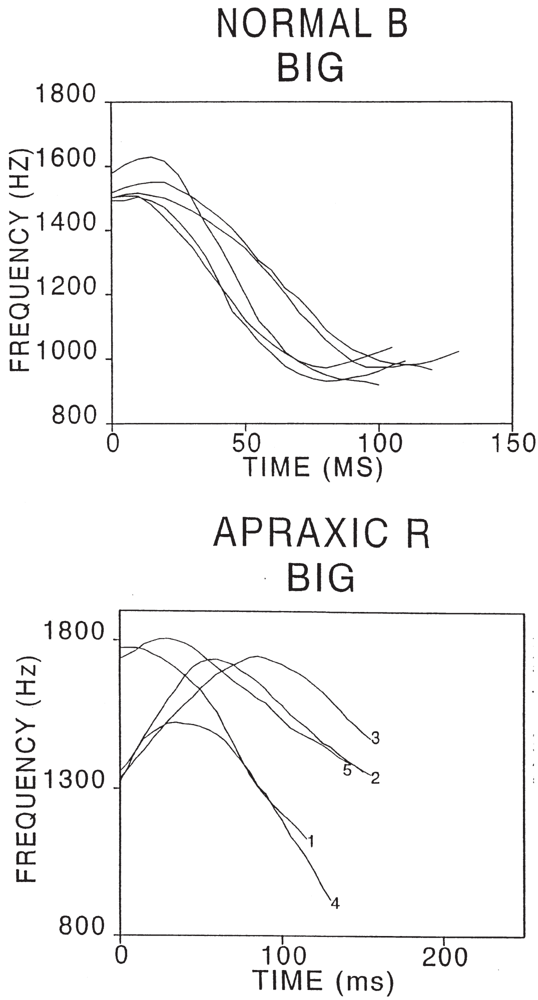


**Figure 7.** F2 trajectories for multiple productions of the word *big* by a normal and apraxic speaker.

damental frequency coupled with an increased noise energy in the region of the 2nd and 3rd formants, may last only 30 to 60 ms and may not be obvious to the listener. Moreover, like the case of the apparently disorganized formant trajectories, these sudden and brief phonatory changes are probably not random losses of control, but appear to be conditioned by phonetic context and characteristics of intonation. A very careful qualitative analysis of this phenomenon should lead to a better idea of a proper quantitative measure. Interestingly, the same kinds of transient voice phenomena are known to occur in normal speech production, but with much more subtle manifestations.

(2) It has been assumed for many years that spirantization, or the production of stop consonants with an incompletely sealed vocal tract, is a signature of Parkinson disease (Logemann & Fisher, 1981). Examination of many records of speech production in a variety of neurological diseases (amyotrophic lateral sclerosis, stroke, cerebellar disease) shows that spirantization occurs frequently in all forms of dysarthria (Weismer, 1997). This kind of observation is important when developing theories of dysarthria designed to explain the origin of intelligibility deficits: the more disease-specific phenomena that are thought to occur, the more treatment strategies may be conceptualized for individual diseases (Duffy, 1995). On the other hand, the identification of a core of speech production symptoms that cuts across a number of disease types may lead to a more general treatment strategy that may be supplemented by the truly disease-specific speech symptoms identified either by quantitative or qualitative analyses.

## ACOUSTIC CORRELATES OF DYSFLUENCY

**Suggested measures:** area under the energy envelope, durations of sounds and pauses, fragmentation, and spectral variations.

**Acoustic-articulatory relationships.** The relationships vary with the type of dysfluency, but the primary acoustic correlate is a disruption of the normal temporal pattern. The disruption can affect the energy envelope, fundamental frequency contour, and spectral properties associated with vowel or consonant segments.

**Acoustic evidence of dysfluency in dysarthria.** Some dysarthric speakers have dysfluencies such as: syllable or word repetitions; sound prolongations; silent blocks or hesitations; and less perceptually obvious features, such as multiple bursts on stop consonant release. These dysfluent behaviors have not been carefully described but have been noted especially for individuals with Parkinson’s disease. Acoustic analyses that hold particular potential for the study of these events include determination of area under the energy envelope (Kuniszyk-Jozkowiak, 1995, 1996) and considerations of sound duration, fragmentation (alternating energy and silence), and spectral properties (Howell, Sackin, & Glenn, 1997; Howell & Wingfield, 1990).

## ASSOCIATIONS OF ACOUSTIC MEASURES WITH PROSODIC AND PARALINGUISTIC ASPECTS

### Prosody

**Suggested measures:** various measures of fundamental frequency (especially intonation), durations of syllables and other units, and intensity.

**Acoustic correlates of prosody.** Prosody is a complex aspect of speech and its correlates can interact with the segmental constituents and paralinguistic properties of an utterance. The three properties of time, fundamental frequency contour, and intensity envelope are typically used separately or in some combination to describe the prosody of speech.

**Acoustic indexes of dysprosody in dysarthria.** Prosodic disturbances are a prominent feature in several forms of dysarthria (Ackermann, Hertrich, & Ziegler, 1993; Darley et al., 1969a, 1969b; Duffy, 1995; Robin, Klouda, & Hug, 1991), but selection of sensitive and efficient procedures has been an obstacle. An immediate problem pertains to the selection of speech material. Several studies indicate that prosodic characteristics differ between passage reading and conversation (see Leuschel & Docherty, 1996, for a review). A general conclusion is that "speech performance in a structured task such as reading, as measured by a range of prosodic parameters, may not be wholly representative of performance in a more naturalistic task such as conversation" (Leuschel & Docherty, p. 164). It appears that conversation may be better than reading in the detection of prosodic abnormalities in dysarthric samples. But the disadvantage of conversation is the lack of control over properties of the utterances, including length, syntactic structure, and phonetic composition. Research by Schlenck, Bettrich, and Willmes (1993) suggests promising directions for analysis based on considerations of tone units and F0 regulation. They report that individuals with severe dysarthria had shorter tone units and higher mean F0 than individuals with mild dysarthria or individuals without neurological disease. Individuals with mild dysarthria had smaller variations of F0 than either severely dysarthric individuals or neurologically normal controls. The variation of prosodic characteristics with severity is an important factor to consider in selecting acoustic measures. Because severely dysarthic individuals tend to produce shorter tone units than mildly impaired individuals, prosodic features such as rhythm or stress pattern may differ between severe and mild dysarthrias simply because of the different lengths of tone unit.

Leuschel and Docherty (1996) took a multi-dimensional stochastic approach to the study of prosody in dysarthria. They collected data for these variables: articulation rate, mean pause duration, number of pauses, articulation/pause time ratio, mean length of utterance, mean utterance duration, mean unstressed vowel duration, percentage of unstressed vowels, intensity range, intensity envelope, F0 mean, F0 range, F0 envelope, and F0 intravowel variation. Data were collected for both reading and conversation, and it was observed that the individuals with dysarthria tended to differ from the neurologically normal subjects more in conversation than in reading. The authors concluded: "Results from both the present and the previous studies have therefore shown that it is important to investigate the performance of dysarthric speakers in different sorts of elicitation tasks, structured and unstructured, in order to arrive at a complete evaluation of individual patients' abilities." (Leuschel & Docherty, p. 166).

## Paralinguistic Variables: Emotion, Tempo, Individual Identity

**Suggested measures:** too numerous to list in detail, but most pertain to fundamental frequency, intensity, and duration.

**Acoustic indexes of paralinguistic features.** The major acoustic properties are the same as for prosody, but the speaking rate is also important.

**Paralinguistic abnormalities in dysarthria.** Most of the acoustic studies published on dysarthria focus on phonetic or prosodic characteristics, which is justifiable because of the immediate concern with intelligibility. Paralinguistic variables have been rather neglected, but this limited research effort should not be interpreted to mean that these characteristics are inconsequential. The emotional and contextual modulation of speech enriches communication, and deficiencies in this capacity can diminish a speaker's effectiveness. Individuals with dysarthria may have reduced abilities to convey emotion when intended, or they may produce patterns that are mistakenly interpreted with respect to emotional valence. In addition, vocal fundamental frequency and speech rate are associated with personality variables such as extroversion, assertiveness, competence, or activity (Brown, Strong, & Rencher, 1974; Brown, Giles, & Thakerar, 1985; Ziegler & Hartmann, 1996). It is likely that many individuals with dysarthria could be perceived as lacking in vitality or liveliness. Studies of the perceptual evaluation of F0 excursions have shown that ratings of liveliness generally vary with power functions of speech rate and the extent of F0 excursions (Traunmuller & Eriksson, 1995). Therefore, dysarthric speech that is slow and monotonous is likely to be rated as having a low degree of liveliness. Studies have demonstrated that listeners do tend to make negative personality assessments of dysarthric speech. Pitcairn, Clemie, Gray, and Pentland (1990) reported that listeners judged speakers with Parkinson's disease to be cold, withdrawn, and anxious.

Speaker recognition has seldom been studied in individuals with dysarthria, but Ziegler and Hartmann (1996) reported some interesting data on listeners' estimation of speaker age for both healthy and dysarthic speakers. Listeners were much poorer at estimating the ages of the speakers with dysarthria. This result may indicate that dysarthric speech is impaired not only in its intelligibility, but also in quality dimensions that interlocutors use in appraising talker characteristics. It would be valuable to know if the neurological disturbances that impair the intelligibility and quality of speech also reduce the distinctiveness of individual speakers. An acoustic property common to many samples of dysarthric speech is a paucity of high-frequency energy. This property helps to explain reduced intelligibility because of the limited information available for many consonants, but it also could account for reduced identifiability of personal characteristics. Although personal characteristics are conveyed through several aspects of the speech signal (Kent & Chial, 1997), certain high-frequency regions may be particularly important (Kitamura & Akagi, 1995).

## TASK-RELATED ACOUSTIC ANALYSES

Acoustic measures can be defined for particular speaking tasks, such as vowel prolongation, syllable repetition, word production, sentence recitation, or conversation. A variety of measures can be proposed for almost any task, and it goes beyond the scope of this paper to consider this material in detail. For general discussion, see Forrest and Weismer (1997), Kent et al., (1997); Liss and Weismer (1992), and Keller, Vigneux, and Laframboise (1991). In addition, commercial systems based on personal computers have been developed for the acoustic analysis of dysarthria. A system that is Macintosh-compatible, DysPhon™ (Keller, 1991), performs analyses for a number of speaking tasks, including sustained sounds, repeated syllables, and sentences. The Motor Speech Profile™ Model 4341 (Kay Elemetrics Corporation) extracts and analyzes a number of parameters, including maximum phonation time and various measures of F0, sound pressure level, diadochokinesis, and F2. Perhaps the simplest tasks for standard analysis are sustained production of a vowel and repetition of a given syllable. Sustained vowels can be conveniently analyzed with a system such as the Multi-Dimensional Voice Program from Kay Elemetrics, or the voice assessment analysis from Dr. Speech Science™ (Tiger Electronics). Syllable repetition at maximal rate (also called alternate motion rate, or diadochokinesis) can be analyzed with various temporal, energy, and spectral measures that define both mean performance and variability of performance within a repetition series. Information on sources for the above systems can be found in Appendix 2.

## ACOUSTIC TYPOLOGIES FOR THE DYSARTHRIAS

Acoustic data can be taken as a composite to describe a particular type of dysarthria (e.g., essential tremor, mixed flaccid-spastic dysarthria in amyotrophic lateral sclerosis, hypokinetic dysarthria in Parkinson’s disease). This application parallels the traditional perceptual assessment of dysarthria in which co-occurring clusters of deviant dimensions are used to support differential diagnosis (Darley et al., 1969a, 1969b; Duffy, 1995). This is a challenging direction for future research, as only limited progress has been made in identifying acoustic dimensions for certain types of dysarthria. It is likely that there will be substantial overlap of the acoustic features for different perceptual types of dysarthria, just as there is overlap of perceptual features. An important goal would be to identify the acoustic features that are most helpful in distinguishing among types and subtypes of dysarthria. To some degree, acoustic correlates can be defined a priori for the primary perceptual features described by Darley et al. (1969a, 1969b). For example, the perceptual feature of monotone in hypokinetic dysarthria is expected to have as its acoustic correlate a limited variation in F0. However, there is no assurance that acoustic properties will al-ways accord with perceptual descriptions. Kim (1994) concluded that listeners' ratings did not distinguish among the three monotonous dimensions of monopitch, monoloudness, and monoduration in samples of speech produced by individuals with Parkinson's disease. She also observed that monopitch and monoloudness were strongly correlated ($r = 0.98$). There are many examples in the literature in which acoustic properties could not be easily reconciled with a perceptual description of the dysarthric sample (Kent, 1996; Kreiman, Gerratt, Kempster, Erman, & Berke, 1993; Kreiman, Gerratt, & Precoda, 1990; Kreiman, Gerratt, Precoda, & Berke, 1992). Acoustic correlates defined a priori from perceptual ratings should be considered as hypotheses only. Neural networks may be an effective means to the identification of efficient classification, using acoustic data, perceptual ratings, or both (Callan & Kent, in press; Leinonen, Kangas, Torkkola, & Juvas, 1992). These networks could help to identify clusters such as those originally described by Darley et al. (1969a, 1969b) for perceptual dimensions. Acoustic methods also hold promise for the detection of subclinical features of neurogenic speech disorders (Parnell & Amerman, 1996; Weismer, 1997).

This work was supported in part by a grant from the NIH-NIDCD, 5 RO1 DC 00319.

## REFERENCES

Ackermann, H., & Hertrich, I. (1994). Speech rate and rhythm in cerebellar dysarthria: an acoustic analysis of syllable timing. *Folia Phoniatrica, 46*, 70–78.

Ackermann, H., Hertrich, I., & Hehr, T. (1995). Oral diadochokinesis in neurological dysarthrias. *Folia Phoniatrica et Logopaedica, 47*, 15–23.

Ackermann, H., Hertrich, I., & Ziegler, W. (1993). Prosodische storungen bei neurologischen Erkrankungen—eine Literaturubersicht. [German] *Fortschritte der Neurologie-Psychiatrie*, *61*, 241–253.

Ackermann, H., & Ziegler, W. (1991a). Cerebellar voice tremor: an acoustic analysis. *Journal of Neurology, Neurosurgery, and Psychiatry, 54*, 74–76.

Ackermann, H., & Ziegler, W. (1991b). Articulatory deficits in Parkinsonian dysarthria: An acoustic analysis. *Journal of Neurology, Neurosurgery, and Psychiatry, 54*, 1093–1098.

Akay, M. (Ed.) (1998). *Time frequency and wavelets in biomedical signal processing.* Piscataway, NJ: IEEE Press.

Allen, G.D. (1978). Vowel duration measurement: A reliability study. *Journal of the Acoustical Society of America, 63*, 1176–1181.

Ansel, B.M., & Kent, R.D. (1992). Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy. *Journal of Speech and Hearing Research, 35,* 296–308.

Aronson, A.E. (1993). *Dysarthria: differential diagnosis* (audio tape). Rochester, MN: Mentor Seminars.

Aronson, A.E., Ramig, L.O., Winholtz, W.S., & Silber, S.R. (1995). Rapid voice tremor or "flutter" in amyotrophic lateral sclerosis. *Annals of Otology, Rhinology, and Laryngology, 101,* 511–518.

Baken, J. (1990). Irregularity of vocal period and amplitude: A first approach to the fractal analysis of voice. *Journal of Voice, 4,* 185–197.

Bielamowicz, S., Kreiman, J., Gerratt, B.R., Dauer, M.S., & Berke, G.S. (1996). Comparison of voice analysis systems for perturbation measurement. *Journal of Speech and Hearing Research, 39,* 126–134.

Blomgren, M., & Robb, M. (1998). How steady are vowel steady-states? *Clinical Linguistics and Phonetics, 12,* 405–415.

Blumstein, S.E., Cooper, W.E., Goodglass, H., Statlender, S., & Gottlieb, J. (1980). Production deficits in aphasia: a voice-onset time analysis. *Brain and Language, 9,* 153–170.

Blumstein, S.E., Cooper, W.E., Zurif, E.B., & Caramazza, A. (1977). The perception and production of voice-onset time in aphasia. *Neuropsychologia, 15,* 371–383.

Bough, I.D., Jr., Heuer, R.J., Sataloff, R.T., Hills, J.R., & Cater, J.R. (1996). Intrasubject variability of objective voice measures. *Journal of Voice, 10,* 166–174.

Boutsen, F.R., Bakker, K., & Duffy, J.R. (1997). Subgroups in ataxic dysarthria. *Journal of Medical Speech-Language Pathology, 5,* 27–36.

Boutsen, F.R., Duffy, J.R., & Aronson, A.E. (1998). Flutter or tremor in hypokinetic dysarthria: A case study. In M.P. Cannito, K.M. Yorkston & D.R. Beukelman (Eds.), *Neuromotor speech disorders: Nature, assessment, and management* (pp. 157–166). Baltimore: Paul H. Brookes.

Brown, B.L., Giles, H., & Thakerar, J.N. (1985). Speaker evaluations as a function of speech rate, accent and context. *Language and Communication, 5,* 207–220.

Brown, B.L., Strong, W.J., & Rencher, A.C. (1974) Fifty-four voices from two: The effects of simultaneous manipulations of rate, mean fundamental frequency and variance of fundamental frequency on ratings of personality from speech. *Journal of the Acoustical Society of America, 55,* 313–318.

Brown, J.R., & Simonson, J. (1963). Organic voice tremor. *Neurology, 13*, 520–525.

Bruce, A. (1996). *Applied wavelets analysis with S-Plus*. New York: Springer.

Buder, E.H. (in press). Acoustic techniques. To appear in R.D. Kent and M.J. Ball (Eds.), *Voice quality measurement*. San Diego: Singular Publishing Group, Inc.

Buder, E.H., Kent, R.D., Kent, J.F., Milenkovic, P., & Workinger, M.S. (1996). FORMOFFA: An automated formant, moment, fundamental frequency, amplitude analysis of normal and disordered speech. *Clinical Linguistics and Phonetics, 10*, 31–54.

Callan, D.E., & Kent, R.D. (in press). The use of self-organizing maps for the classification of voice disorders. To appear in R.D. Kent and M.J. Ball (Eds.), *Voice quality measurement*. San Diego: Singular Publishing Group, Inc.

Caruso, A.J., & Burton, E.K. (1987). Temporal acoustic measures of dysarthria associated with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research, 30*, 80–87.

Chen, M.Y. (1997). Acoustic correlates of English and French nasalized vowels. *Journal of the Acoustical Society of America, 102*, 2360–2370.

Cisneros, E., & Braun, C.M. (1995). [Vocal and respiratory diadochokinesis in Friedreich's ataxia. Neuropathological correlations.] [French] *Revue Neurologique, 151*, 113–123.

Collins, M. (1984). Integrating perceptual and instrumental procedures in dysarthria assessment. *Journal of Communication Disorders, 5*, 159–170.

Darley, F.L., Aronson, A.E., & Brown, J.R. (1969a). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12*, 249–269.

Darley, F.L., Aronson, A.E., & Brown, J.R. (1969b). Cluster of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research, 12*, 462–496.

Darley, F.L., Aronson, A.E., & Brown, J.R. (1975). *Motor Speech Disorders*. Philadelphia: W.B. Saunders.

Dejonckere, P.H., & Wieneke, G.H. (1992). Cepstrum analysis of pathological voices: Correlations with other acoustic parameters and with perceptual scaling. In *The new frontiers of oto-rhino-laryngology in Europe, IV*, 75–78. Proceedings, 2nd European Congress of Oto-rhino-larynology and Cervico-facial Surgery. Bologna: Monduzzi.

Dejonckere, P. H., & Wieneke, G. H. (1993). Cepstra of normal and pathological voices, in correlation with acoustic, aerodynamic, and perceptual data.

In M.J. Ball & M. Duckworth (Eds.), *Advances in clinical phonetics* (pp. 217–226). Amsterdam: John Benjamins.

Dejonckere, P.H., & Wieneke, G.H. (1994). Spectral, cepstral, and aperiodicity characteristics of pathological voices before and after phonosurgical treatment. *Clinical Linguistics and Phonetics, 8,* 161–169.

Deller, J.R., Jr., Lilu, M.S., Ferrier, L.J., & Robichaud, P. (1993). The Whitaker database of dysarthric (cerebral palsy) speech. *Journal of the Acoustical Society of America*, *93*, 3516–3518.

Doherty, E.T., & Shipp, T. (1988). Tape recorder effects on jitter and shimmer extraction. *Journal of Speech and Hearing Research, 31*, 485–490.

Duffy, J.R. (1995). *Motor speech disorders: Substrates, differential diagnosis, and management.* St. Louis: Mosby-YearBook, Inc.

Farmer, A. (1980). Voice onset time production in cerebral palsied speakers. *Folia Phoniatrica, 32*, 267–273.

Forrest, K., & Weismer, G. (1997). Acoustic analysis of dysarthric speech. In M.R. McNeil (Ed.), *Clinical management of sensorimotor speech disorders* (pp. 63–80). New York: Thieme.

Freeman, F.J., Sands, E.S., & Harris, K.S. (1978). Temporal coordination of phonation and articulation in a case of verbal apraxia: A voice onset time study. *Brain and Language*, *6,* 106–111.

Gath, I., & Yair, E. (1988). Analysis of vocal tract parameters in Parkinsonian speech. *Journal of the Acoustical Society of America, 84*, 1628–1634.

Gentil, M. (1990). Acoustic characteristics of speech in Friedreich's disease. *Folia Phoniatrica, 42*, 125–134.

Gerratt, B. (1983). Formant frequency fluctuation as an index of motor steadiness in the vocal tract. *Journal of Speech and Hearing Research, 26*, 297–304.

Gopal, H.S. (1995). Technical issues underlying the development and use of a speech research laboratory. In A. Syrdal, R. Bennett & S. Greenspan (Eds.), *Applied speech technology* (pp. 315–342). Boca Raton: CRC Press.

Green, J.R., Buder, E.H., Rodda, P.R., & Moore, C.A. (1998). Reliability of measurements across several acoustic voice analysis systems. In M.P. Cannito, K.M. Yorkston & D.R. Beukelman (Eds.), *Neuromotor speech disorders: Nature, assessment, and management* (pp. 275–292). Baltimore: Paul H. Brookes Publishing Company.

Harmes, S., Daniloff, R., Hoffman, P., Lewis, P., Kramer, M., & Absher, R. (1984). Temporal and articulatory control of fricative articulation by speakers with Broca's aphasia. *Journal of Phonetics*, *12*, 367–385.

Hartelius, L., Buder, E.H., & Strand, E.A. (1997). Long-term phonatory instability in individuals with multiple sclerosis. *Journal of Speech, Language, and Hearing Research, 40,* 1056–1072.

Hertrich, I., & Ackermann, H. (1993). Acoustic analysis of speech prosody in Huntington's and Parkinson's disease—a preliminary report. *Clinical Linguistics and Phonetics, 7,* 285–297.

Hertrich, I., & Ackermann, H. (1995). Gender-specific vocal dysfunctions in Parkinson's disease. *Annals of Otology Rhinology & Laryngology, 104,* 197–202.

Hertrich, I., & Ackermann, H. (1997). Acoustic analysis of durational speech parameters in neurological dysarthrias. In Y. Lubrun (Ed.), *From the brain to the mouth* (pp. 11–48). Dordrecht, The Netherlands: Kluwer Academic Publishers.

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

Herzel, H., Berry, D., Titze, I., & Steinecke, I. (1995). Nonlinear dynamics of the voice: Signal analysis and biomechanical modeling. *Chaos, 5,* 30–34.

Hess, W.J. (1992). Pitch and voicing determination. In S. Furui & M.M. Sondhi (Eds.), *Advances in speech signal processing* (pp. 3–48). New York: Marcel Dekker, Inc.

Hiraoka, N., Kitazoe, Y., & Ueta, H. (1984). Harmonic-intensity analysis of normal and hoarse voices. *Journal of the Acoustical Society of America, 76,* 1648–1651.

Howell, P., Sackin, S., & Glenn, K. (1997). Development of a two-stage procedure for the automatic recognition of dysfluencies in the speech of children who stutter: II. ANN recognition of repetitions and prolongations with supplied word segment markers. *Journal of Speech and Hearing Research, 40*, 1085–1096.

Howell, P., & Wingfield, T. (1990). Perceptual and acoustic evidence for reduced fluency in the vicinity of stuttering episodes. *Language and Speech, 33*, 31–46.

Hudson, A., & Holbrook, A. (1981). A study of the reading fundamental vocal frequency of young black adults. *Journal of Speech and Hearing Research, 24*, 197–201.

Hudson, A., & Holbrook, A. (1982). Fundamental frequency characteristics of young black adults: Spontaneous speaking and oral reading. *Journal of Speech and Hearing Research, 25,* 25–28.

Ingrisano, D.R.-S., Perry, C.K., & Jepson, K.R. (1998). Environmental noise: A threat to automatic voice analysis. *American Journal of Speech-Language Pathology, 7*, 91–96.

Itoh, M., Sasanuma, S., Tatsumi, I., Murakami, S., Fukusako, Y., & Suzuki, T. (1982). Voice onset time characteristics in apraxia of speech. *Brain and Language*, *17*, 193–210.

Jafari, M., Till, J.A., Truesdell, L.F., & Law-Till, C.B. (1993). Time-shift, trial, and gender effects on vocal perturbation measures. *Journal of Voice, 7*, 326–336.

Karnell, M.P. (1991). Laryngeal perturbation analysis: Minimum length of analysis window. *Journal of Speech and Hearing Research, 34*, 544–548.

Karnell, M.P., Hall, K.D., & Landahl, K.L. (1995). Comparison of fundamental frequency and perturbation measurements among three analysis systems. *Journal of Voice*, *9*, 383–393.

Keller, E. (1991). DysPhon (Version 2.0). [Computer software]. Neuchatel, Switzerland: InfoSignal, Inc.

Keller, E., Vigneux, P., & Laframboise, M. (1991). Acoustic analysis of neurologically impaired speech. *British Journal of Disorders of Communication, 26*, 75–94.

Kent, R.D. (1979). Isovowel lines for evaluating vowel formant structures in speech disorders. *Journal of Speech and Hearing Disorders, 44*, 513–521.

Kent, R.D. (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology*, *7*, 7–23.

Kent, R.D., & Chial, M.R. (1997). Talker identification. In D.L. Faigman, D. Kaye, M.J. Saks & J. Sanders (Eds.), *Modern scientific evidence: The Law and science of expert testimony*. St. Paul: West Publishing Co.

Kent, R.D., Kent, J.F., Rosenbek, J.C., Vorperian, H.K., & Weismer, G. (1997). A speaking task analysis of the dysarthria in cerebellar disease. *Folia Phoniatrica et Logopaedica, 49*, 63–82.

Kent, R.D., Kent J.F., Weismer, G., Sufit, R.L., Rosenbek, J.C., Martin, R.E., & Brooks, B.R. (1990). Impairment of speech intelligibility in men with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Disorders, 55*, 721–729.

Kent, R.D., Kim, H., Weismer, G., Kent, J.F., Rosenbek, J.C., Brooks, B.R., & Workinger, M. (1994). Laryngeal dysfunction in neurological disease:

amyotrophic lateral sclerosis, Parkinson's disease, and stroke. *Journal of Medical Speech-Language Pathology, 2,* 157–175.

Kent, R.D., Netsell, R., & Abbs, J. (1979). Acoustic characteristics of dysarthria associated with cerebellar disease. *Journal of Speech and Hearing Research, 22,* 627–648.

Kent, R.D., & Read, W.C. (1992). *The Acoustic Analysis of Speech.* San Diego: Singular Publishing Group.

Kent, R.D., Sufit, R.L., Rosenbek, J.C., Kent, J.F., Weismer, G., Martin, R.E., & Brooks, B.R. (1991). Speech deterioration in amyotrophic lateral sclerosis: A case study. *Journal of Speech and Hearing Research, 34,* 1269–1275.

Kim, H.-H. (1994). Monotony of speech production in Parkinson's disease: Acoustic entities and their perceptual relations. Unpublished Ph.D. dissertation, University of Wisconsin-Madison.

King, J.B., Ramig, L.O., Lemke, J.H., & Horii, Y. (1994). Parkinson's disease: Longitudinal changes in acoustic parameters of phonation. *Journal of Medical Speech-Language Pathology, 2,* 29–42.

Kitamura, T., & Akagi, M. (1995). Speaker individualities in speech spectral envelopes. *Journal of the Acoustical Society of Japan, 16,* 283–289.

Klatt, D.H. (1975). Voice onset time, frication, and aspiration in word initial consonant clusters. *Journal of Speech and Hearing Research, 18,* 686–705.

Kojima, Y., Shimoyama, I., Ninchoij, T., & Uemura, K. (1989). [Significance of "on-off phonation test" to detect oral diadochokinesia in dysarthric patients.] [Japanese] *No to Shinkei, 41,* 1091–1095.

Kreiman, J., Gerratt, B.R., Kempster, G.B., Erman, A., & Berke, G. (1993). Perceptual evaluation of voice quality: Review, tutorial and a framework for future research. *Journal of Speech and Hearing Research, 36,* 21–40.

Kreiman, J., Gerratt, B.R., & Precoda, K. (1990). Listener experience and perception of voice quality. *Journal of Speech and Hearing Research, 33,* 103–115.

Kreiman, J., Gerratt, B.R., Precoda, K., & Berke, G.S. (1992). Individual differences in voice quality perception. *Journal of Speech and Hearing Research, 35,* 512–520.

Kuniszyk-Jozkowiak, W. (1995). The statistical analysis of speech envelopes in stutterers and nonstutterers. *Journal of Fluency Disorders, 20,* 11–23.

Kuniszyk-Jozkowiak, W. (1996). A comparison of speech envelopes of stutterers and nonstutterers. *Journal of the Acoustical Society of America, 100,* 1105–1110.

Leek, M.R. (1995). Will a good disc last? *CD-ROM Professional, 8*, 102–110.

Leinonen, L., Kangas, J., Torkkola, K., & Juvas, A. (1992). Dysphonia detected by pattern recognition of spectral composition. *Journal of Speech and Hearing Research*, *35*, 287–295.

Leuschel, A., & Docherty, G.J. (1996). Prosodic assessment of dysarthria. In D.A. Robin, K.M. Yorkston & D.R. Beukelman (Eds.), *Disorders of motor speech: Assessment, treatment, and clinical characterization* (pp. 155–178). Baltimore: Paul H. Brookes Publishing Company.

Linville, S.E. (1996). The sound of senescence. *Journal of Voice, 10,* 190–200.

Lisker, L., & Abramson, A.S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word, 20,* 384–422.

Liss, J.M., & Weismer, G. (1992). Qualitative acoustic analysis in the study of motor speech disorders. *Journal of the Acoustical Society of America, 92,* 2984–2987.

Logemann, J.A., & Fisher, H.B. (1981). Vocal tract control in Parkinson's disease: Phonetic feature analysis of misarticulations. *Journal of Speech and Hearing Research. 46,* 348–352.

Ludlow, C.L., Bassich, C.J., Connor, N.P., & Coulter, D.C. (1984). Phonatory characteristics of vocal fold tremor. *Journal of Phonetics, 14,* 509–515.

Ludlow, C., Coulter, D. & Gentges, F. (1983). The differential sensitivity of frequency perturbation to laryngeal neoplasms and neuropathologies. In D. Bless & J.H. Abbs (Eds.), *Vocal fold physiology: Contemporary research and clinical issues* (pp. 381–392). San Diego: College-Hill Press.

Mende, W., Herzel, H., & Wermke, K. (1990). Bifurcations and chaos in newborn infant cries. *Physics Letters, 145,* 418–424.

Metz, D.E., Conture, E.G., & Caruso, A.J. (1979). Voice onset time, frication and aspiration during stutterers' fluent speech. *Journal of Speech and Hearing Research, 22*, 649–656.

Monsen, R.B., & Engebretson, A.M. (1983). The accuracy of formant frequency measurements: A comparison of spectrographic analysis and linear prediction. *Journal of Speech and Hearing Research*, *26*, 89–97.

Murdoch, B.E., Thompson, E.C., & Stokes, P.D. (1994). Phonatory and laryngeal dysfunction following upper motor neuron vascular lesions. *Journal of Medical Speech-Language Pathology, 2,* 177–189.

Nieman, G.S., Klich, R.J., & Shuey, E.M. (1983). Voice onset time in young and 70-year-old women. *Journal of Speech and Hearing Research, 26*, 118–123.

Parnell, M.M., & Amerman, J.D. (1996). An 11-year follow-up of motor speech abilities in elderly adults. *Clinical Linguistics and Phonetics, 10*, 103–118.

Perry, C., Ingrisano, D., & Blair, W.B. (1996). The influence of recording systems on jitter and shimmer estimates. *American Journal of Speech-Language Pathology, 5*, 86–90.

Philippbar, S.A., Robin, D.A., & Luschei, E.S. (1989). Limb, jaw, and vocal tremor in Parkinson's patients. In K.M. Yorkston & D.R. Beukelman (Eds.), *Recent advances in clinical dysarthria* (pp. 165–197). Boston: College-Hill Press.

Pirello, K., Blumstein, S.E., & Kurowski, K. (1997). The characteristics of voicing in syllable-initial fricatives in American English. *Journal of the Acoustical Society of America, 101*, 3754–3765.

Pitcairn, T.K., Clemie, S., Gray, J.M., & Pentland, B. (1990). Non-verbal cues in the self-presentation of Parkinsonian patients. *British Journal of Clinical Psychology, 29*, 177–184.

Plante, F., Berger-Vachon, C., & Kauffman, I. (1993). Acoustic discrimination of velar impairment in children. *Folia Phoniatrica, 45*, 112–119.

Portnoy, R.A., & Aronson, A.E. (1982). Diadochokinetic syllable rate and regularity in normal and in spastic and ataxic dysarthric subjects. *Journal of Speech and Hearing Disorders, 47*, 324–328.

Ramig, L.A., Scherer, R.C., Titze, I.R., & Ringel, S.P. (1988). Acoustic analysis of voices of patients with neurological disease: Rationale and preliminary data. *Annals of Otology, Rhinology, and Laryngology, 97*, 64–172.

Ramig, L.A., & Shipp, T. (1987). Comparative measures of vocal tremor and vocal vibrato. *Journal of Voice, 2*, 162–167.

Read, C., Buder, E.H., & Kent, R.D. (1992). Speech analysis systems: An evaluation. *Journal of Speech and Hearing Research, 35*, 314–332.

Robb, M., Blomgren, M., & Chen, Y. (1998). Formant frequency fluctuation in stuttering and nonstuttering adults. *Journal of Fluency Disorders, 23*, 73–84.

Robin, D.A., Klouda, G., & Hug, L. (1991). Neurogenic disorders of prosody. In M.P. Cannito & D. Vogel (Eds.), *Treating disordered speech motor control: For clinicians by clinicians* (pp. 241–271). Austin: Pro-Ed.

Ryalls, J., Zipprer, A. & Baldauff, P. (1997). A preliminary investigation of the effects of gender and race on voice onset time. *Journal of Speech, Language, and Hearing Research, 40*, 642–645.

Schlenck, K.-J., Bettrich, R., & Willmes, Z.K. (1993). Aspects of disturbed prosody in dysarthria. *Clinical Linguistics and Phonetics, 7*, 119–128.

Shinn, P., & Blumstein, S.E. (1983). Phonetic deterioration in aphasia: Acoustic analysis of spectral characteristics for place of articulation. *Brain & Language, 20*, 90–114.

Smith, B.L., Hillenbrand, J., & Ingrisano, D. (1986). A comparison of temporal measures of speech using spectrograms and digital oscillograms. *Journal of Speech and Hearing Research, 29*, 270–274.

Speliotis, D.E., & Peter, K.J. (1991). Corrosion study of metal particle, metal film, and BA-ferrite tape. *IEEE Transactions on Magnetics, 27*, 4724–4726.

Steinecke, I., & Herzel, H. (1995). Bifurcations in an asymmetric vocal-fold model. *Journal of the Acoustical Society of America, 97*, 1874–1884.

Strand, E.A., Buder, E.H., Yorkston, K.M., & Ramig, L.O. (1994). Differential phonatory characteristics of four women with amyotrophic lateral sclerosis. *Journal of Voice, 8*, 327–339.

Strand, E.A., & Yorkston, K.M. (1994). Description and classification of individuals with dysarthria: a 10-year review. In J.A. Till, K.M. Yorkston & D.R. Beukelman (Eds.), *Motor speech disorders: Advances in assessment and treatment* (pp. 37–54). Baltimore: Paul H. Brookes.

Swartz, B. (1992). Gender differences in voice onset time. *Perceptual and Motor Skills, 75*, 983–992.

Syrdal, A.K., & Gopal, H.S. (1986). A perceptual model of vowel representation of American English vowels. *Journal of the Acoustical Society of America, 79*, 1086–1100.

Titze, I.R. (1994). Toward standards in acoustic analysis of voice. *Journal of Voice, 8*, 1–7.

Titze, I.R., & Liang, H. (1993). Comparison of $F_0$ extraction methods for high-precision voice perturbation measurements. *Journal of Speech and Hearing Research, 36*, 1120–1133.

Titze, I.R., & Winholtz, W.S. (1993). Effect of microphone type and placement on voice perturbation measurements. *Journal of Speech and Hearing Research, 36*, 1177–1190.

Tjaden, K., & Turner, G.S. (1997). Spectral properties of fricatives in amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research, 40*, 1358–1372.

Traunmuller, H., & Eriksson, A. (1995). The perceptual evaluation of $F_0$ excursions in speech as evidenced in liveliness estimations. *Journal of the Acoustical Society of America, 97*, 1905–1915.

Turner, G.S., Tjaden, K., & Weismer, G. (1995). The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research, 38*, 1001–1013.

Venkatagiri, H.S. (1996). The quality of digitized and synthesized speech: What clinicians should know. *American Journal of Speech-Language Pathology, 5*, 31–42.

Walton, J., & Orlikoff, R. (1994). Speaker race identification from acoustic cues in the vocal signal. *Journal of Speech and Hearing Research, 37*, 738–745.

Watanabe, S., Arasaki, K., Nagata, H., & Shouji, S. (1994). [Analysis of dysarthria in amyotrophic lateral sclerosis—MRI of the tongue and formant analysis of vowels.] [Japanese]. *Rinsho Shinkeigaku [Clinical Neurology], 34*, 217–223.

Watson B., & Alfonso, P.J. (1982). Comparison of LRT and VOT values between stutterers and nonstutterers. *Journal of Fluency Disorders, 7*, 219–242.

Weismer, G. (1979). Sensitivity of voice-onset time (VOT) to certain segmental features in speech production. *Journal of Phonetics, 7*, 197–204.

Weismer, G. (1984). Articulatory characteristics of Parkinsonian dysarthria: Segmental and phrase-level timing, spirantization, and glottal-supraglottal coordination. In M.R. McNeil, J.C. Rosenbek & A.E. Aronson (Eds.), *The dysarthrias: Physiology, acoustics, perception, management* (pp. 101–130). San Diego: College-Hill Press.

Weismer, G. (1997). Motor speech disorders. In W.J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp.191–219). Cambridge, MA: Blackwell.

Weismer, G., & Fromm, D. (1983). Acoustic analysis of geriatric utterances: Segmental and non-segmental characteristics which relate to laryngeal function. In D.M. Bless and J.H. Abbs (Eds.), *Vocal Fold Physiology: Contemporary Research and Clinical Issues* (pp. 317–332). San Diego: College-Hill.

Weismer G., & Liss J.M. (1991). Age and speech motor control. In D. Ripich (Ed.), *Handbook of Aging and Communication* (pp. 205–226). Austin, TX: PRO-ED.

Weismer, G., & Martin, R.E. (1992). Acoustic and perceptual approaches to the study of intelligibility. In R.D. Kent (Ed.) *Intelligibility in speech disorders: Theory, measurement and management* (pp. 68–118). Amsterdam: John Benjamins.

Weismer, G., Martin, R.E., Kent, R.D., & Kent, J.F. (1992). Formant trajectory characteristics of males with amyotrophic lateral sclerosis. *Journal of the Acoustical Society of America, 91*, 1085–1098.

Weismer, G., Tjaden, K., & Kent, R.D. (1995). Can articulatory behavior in speech disorders be accounted for by theories of normal speech production? *Journal of Phonetics, 23*, 149–164.

Winholtz, W.S., & Ramig, L.O. (1992). Vocal tremor analysis with the vocal demodulator. *Journal of Speech and Hearing Research, 35*, 562–573.

Winholtz, W.S., & Titze, I.R. (1997). Miniature head-mounted microphone for voice perturbation analysis. *Journal of Speech, Language, and Hearing Research, 40*, 894–899.

Wornell, G. (1996). *Signal processing with fractals: A wavelet-based approach.* Upper Saddle River: Prentice Hall.

Ziegler, W. (1987). Phonetic realization of phonological contrast in aphasic patients. In J.H. Ryalls (Ed.), *Phonetic approaches to speech production in aphasia and related disorders* (pp. 163–197). Boston: Little, Brown and Co.

Ziegler, W., & Hartmann, E. (1996). Perceptual and acoustic methods in the evaluation of dysarthric speech. In M.J. Ball & M. Duckworth (Eds.), *Advances in clinical phonetics* (pp. 91–114). Amsterdam: John Benjamins.

Ziegler, W., & von Cramon, D. (1983a). Vowel distortion in traumatic dysarthria: A formant study. *Phonetica, 40*, 63–78.

Ziegler, W., & von Cramon, D. (1983b). Vowel distortion in traumatic dysarthria: Lip rounding versus tongue advancement. *Phonetica, 40*, 312–322.

Ziegler, W., & von Cramon, D. (1986). Spastic dysarthria after acquired brain injury: An acoustic study. *British Journal of Disorders of Communication, 21*, 173–187.

Ziegler, W., & Wessel, K. (1996). Speech timing in ataxic disorders. *Neurology, 47*, 208–214.

Zwirner, P., Murry, T., & Woodson, G. E. (1991). Phonatory function of neurologically impaired patients. *Journal of Communication Disorders, 24*, 287–300.

Zyski, B.J., & Weisiger, B.E. (1987). Identification of dysarthia types based on perceptual analysis. *Journal of Communication Disorders, 20*, 367–378.

## CONTINUING EDUCATION

## Acoustic Studies of Dysarthric Speech: Methods, Progress, and Potential

## SUGGESTED READINGS

Forrest, K., & Weismer, G. (1997). Acoustic analysis of dysarthric speech. In M.R. McNeil (Ed.), *Clinical management of sensorimotor speech disorders* (pp. 63–80). New York: Thieme.

Keller, E., Vigneux, P., & Laframboise, M. (1991). Acoustic analysis of neurologically impaired speech. *British Journal of Disorders of Communication*, *26*, 75–94.

Kent, R.D. (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology*, *7*, 7–23.

Venkatagiri, H.S. (1996). The quality of digitized and synthesized speech: What clinicians should know. *American Journal of Speech-Language Pathology*, *5*, 31–42.

Weismer, G. (1997). Motor speech disorders. In W.J. Hardcastle, & J. Laver (Eds.), *The handbook of phonetic sciences* (pp.191–219). Cambridge: Blackwell.

QUESTIONS

1. Which of the following is *not* considered a frequency-domain analysis of sound?
   a. Fourier spectrum
   b. Linear predictive coding spectrum
   c. Cepstrum
   d. Waveform

2. What is the most likely articulatory correlate for the acoustic cue of increasing separation in frequency between the first formant (F1) and second formant (F2)?
   a. Forward (anterior) movement of the tongue
   b. Upward (superior) movement of the tongue
   c. Both a and b
   d. Movement of the tongue to a central position

3. What is the best term for the composite measure derived by summing the values for closure duration and voice onset time for a voiceless stop occurring between voiced elements (e.g., the [*k*] in the word "taking?"
   a. Syllable duration

b. Laryngeal devoicing gesture
c. Stop gap
d. Frication duration

4. The presence of high-frequency noise energy during a stop gap is acoustic evidence for which of the following?
   a. Voicing (vocal fold vibrations)
   b. Nasalization
   c. Spirantization
   d. All of the above

5. If a speaker with severe dysarthria has greatly reduced tongue movements for vowel articulation, which of the following is most useful as an acoustic correlate of this articulatory limitation?
   a. Slope of F2 transitions
   b. Range of F1 frequencies
   c. Area of vowel quadrilateral in F1–F2 plane
   d. Bandwidth of vowel formants

6. Sometimes a distinction is made between wow, tremor, and flutter in analyses of rhythmic or oscillatory behavior. The term “flutter” is used especially for oscillations of
   a. Fewer than 3 Hz
   b. Between 3 and 7 Hz
   c. Between 6 and 12 Hz
   d. Between 10 and 20 Hz

7. Which of the following units of analysis is most suitable for prosodic analysis?
   a. Laryngeal devoicing gesture
   b. Formant transition
   c. Frication duration
   d. Tone unit

8. Place of articulation for stops and fricatives is signaled especially by which acoustic cues?
   a. First-formant (F1) frequency and voice bar
   b. Spectrum of noise and pattern of second-formant (F2) and third-formant (F3) transitions
   c. Spectrum of aspiration or frication noise and voice onset time
   d. Duration of noise energy and duration of following vowel

9. If the goal of analysis is to determine the formant structure for vowels, which methods are most suitable?
   a. FFT spectra
   b. Waveform and energy envelope

c. LPC spectra
d. Both a and c

10. In studies of vowel durations in normal adult speech, the estimated reliability is about
a. 1 ms
b. 20 ms
c. 50 ms
d. 100 ms

## APPENDIX 1: GLOSSARY

*Note*: many items are reprinted with permission from R.D. Kent and C. Read, *The Acoustic Analysis of Speech* (1992).

*Antiformant*—a property of the vocal tract transfer function in which energy is not passed effectively through the system; opposite in effect to a formant. Antiformants, or zeros, arise because of divided passages or constrictions in the vocal tract.

*Bandwidth*—a measure of the frequency band of a sound, especially a resonance. Conventionally, bandwidth is determined at the half-power ("3 db down") points of the frequency response curve. That is, both the lower and higher frequencies that define the bandwidth are 3 db less intense than the peak energy in the band.

*Bark scale*—a nonlinear transformation of frequency that is thought to correspond to the analysis accomplished by the ear. The Bark scale is closely related to the concept of critical band in auditory perception.

*Cepstrum*—A Fourier transform of the power spectrum of a signal. The transform is described in terms of *quefrency* (note the transliteration from frequency), which has time-like properties. The cepstrum is used to determine the fundamental frequency of a speech signal. Voiced speech tends to have a strong cepstral peak, at the first *rahmonic* (note transliteration from harmonic).

*Coupling*—interaction between two or more systems; e.g., oral-nasal coupling refers to the degree of interaction between the two resonating cavities. No coupling means no interaction.

*Damping*—the rate of absorption of sound energy; related to bandwidth.

*Fast Fourier transform (FFT)*—an algorithm commonly used in microcomputer programs to calculate a Fourier spectrum. The FFT is a special type of DFT in which the number of points transformed is a power of 2. The number of points expresses the bandwidth of analysis; the higher the value, the narrower the bandwidth.

*Filter*—a hardware device or software program that provides a frequency-dependent transmission of energy. Commonly, a filter is used to exclude energy at certain frequencies while passing the energy at other frequencies. A low-pass filter passes the frequencies below a certain cut-off frequency; a high-pass filter passes the frequencies above a certain cut-off frequency; and a band-pass filter passes the energy between a lower and upper cut-off frequencies.

*Formant*—a resonance of the vocal tract. A formant is specified by its center frequency (commonly called formant frequency) and bandwidth. Formants are denoted by integers that increase with the relative frequency location of the formants. F1 is the lowest-frequency formant, F2 is the next highest, and so on.

*Fourier transform*—a mathematical procedure that converts a series of values in the time domain (waveform) to a set of values in the frequency domain (spectrum). The spectrum is the Fourier transform of a waveform; the waveform is the inverse Fourier transform of the spectrum.

*Formant transition*—a change in formant pattern, typically associated with a phonetic boundary; for example, the CV formant transition refers to formant pattern changes associated with the consonant-vowel transition.

*Frequency-domain operation*—an operation that is performed in the frequency domain, e.g., with a FFT or LPC spectrum.

*Fundamental frequency*—The lowest frequency (first harmonic) of a periodic signal. In speech, the fundamental frequency refers to the first harmonic of the voice. Fundamental frequency is the reciprocal of the fundamental period. Ideally, fundamental frequency is used to refer to a *physical* measure of the lowest periodic component of vocal fold vibration. *Pitch* should be used to indicate the perceptual phenomenon in which stimuli can be rated along a continuum of low to high. See *Pitch Determination Algorithm*.

*Harmonic*—an integer multiple of the fundamental frequency in voiced sounds. Ideally, the voice source can be conceptualized as a line spectrum in which energy appears as a series of harmonics.

*Laryngeal devoicing gesture*—for voiceless stops flanked by voiced segments, the sum of the closure duration and the voice onset time.

*Linear predictive coding (LCP)*—a class of methods used to obtain a spectrum. Linear predictive coding uses a weighted linear sum of samples to predict an upcoming value.

*Narrow-band analysis*—an analysis in which the analyzing bandwidth is relatively narrow (such as 45 Hz in speech analysis). A narrow-band analysis is preferred when the interest is to increase frequency resolution, as in the analysis of harmonics for a man's voice.

*Nasal formant*—the low-frequency resonance associated with the nasal tract. For men's speech, the nasal formant has a frequency of less than 500 Hz.

*Normalization*—a correction for variance. *Speaker normalization* refers to the correction or scaling that reduces variability in acoustic measures such as formant frequencies. *Time normalization* refers to the correction or scaling that reduces variability in the durations of sound sequences.

*Nyquist sampling theorem*—this theorem states that a digital representation requires at least two sampling points for every periodic cycle in the signal of interest. Therefore, the sampling rate of digitization should be at least twice the highest frequency of interest in the signal to be analyzed. Unfortunately, the term *Nyquist Frequency* is inconsistently used. Some use it to indicate the highest frequency of interest in an analysis; others use it to refer to twice the highest frequency of interest, i.e. to the sampling rate needed to prevent aliasing.

*Pitch determination algorithm (PDA) (*also *pitch extraction)*—a procedure used to extract the fundamental frequency of a speech signal. Although the term *pitch* strictly should be used to refer to a perceptual phenomenon, it is often used in speech analysis to refer to fundamental frequency.

*Preemphasis*—in speech analysis, a filtering that boosts high-frequency energy relative to low-frequency energy. Because speech normally contains its strongest energy in the low frequencies, these frequencies would dominate analysis results if preemphasis were not performed.

*Prevoicing*—the onset of voicing before the appearance of a supraglottal articulatory event; e.g., for stops, prevoicing means that voicing precedes the stop release. Also called voicing lead.

*Radiation characteristic*—the term in source-filter theory associated with the radiation of sound from the lips to the atmosphere. It is typically expressed as a 6 dB per octave increase in sound energy (hence, a high-pass filter).

*Reynold's number*—a dimensionless number that serves as an index of the development of turbulence.

*RMS amplitude*—root-mean-square measurement of signal amplitude, the abbreviation *rms* is based on the fundamental operations of squaring the individual values, taking their mean, and obtaining the square root of the mean.

*Sampling theorem*—this theorem, developed by Nyquist, states that S samples per second are needed to represent a waveform with a bandwidth of S/2 Hz.

*Segmentation*—the delineation of successive sound segments in a speech signal. Typically, segmentation yields units such as phonemes, allophones, or some other phonetic segment.

*Source-filter theory*—a theory of the acoustic production of speech that states that the energy from a sound source is modified by a filter or set of filters; for example, for vowels the vibrating vocal folds are usually the source of sound energy, and the vocal tract resonances (formants) are the filters.

*Spectrogram*—a pattern for sound analysis containing information on intensity, frequency and time. The typical spectrogram provides a three-dimensional display of time on the horizontal axis), frequency on the vertical axis, and intensity on the gray scale. A spectrogram can be printed as hard copy or displayed on a video monitor.

*Spectrum*—a graph showing the distribution of signal energy as a function of frequency; a plot of intensity by frequency.

*Stop gap*—the acoustic interval corresponding to articulatory closure for a stop or affricate consonant; it is identified on a spectrogram as an interval of relatively low energy, conspicuously lacking in formant pattern or noise.

*Time-domain operation*—an operation that is performed in the time domain, for example, calculations performed with respect to the waveform of a sound.

*Voice bar*—a band of energy, typically reflecting the first harmonic of the voice source, that appears on a spectrogram; it is indicative of voicing.

*Voice onset time (VOT)*—a measure of the time between a supraglottal event and the onset of voicing; for stops, VOT is the interval between release of the stop (usually determined acoustically as the stop burst) and the appearance of periodic modulation (voicing) for a following sound.

*Waveform*—a graph showing the amplitude versus time function for a continuous signal such as the acoustic signal of speech.

*Wavelength*—the distance that a periodic sound travels in one complete cycle. Wavelength = speed of sound/frequency.

*Wide-band analysis*—an analysis in which a relatively large analyzing bandwidth is used (such as 300 Hz in speech analysis). A wide-band analysis is preferred when the primary concern is to reveal formant pattern or to increase time resolution.

*Wow*—in voice analysis refers to a low-frequency vocal tremor.

## Appendix 2: Commercial Systems of the Acoustic Analysis of Disordered Speech

*Note:* Systems are listed by manufacturer or supplier.

**Source:** InfoSignal Inc., Rue de la Dime 80, Ch-2000, Neuchatel, Switzerland:

Dysphon™ is a set of computer programs for the acoustic analysis of dysarthria including: speech examination; Signalyze™ (a signal analysis program for the Macintosh personal computer), speech evaluation (task scoring), and DysPhon-ExSys I (an expert system to automatize the evaluation).

**Source:** Kay Elemetrics Corporation, 2 Bridgewater Lane, Lincoln Park, NJ 07035-1488:

Motor Speech Profile (SMP)™, Model 4341, extracts and analyzes speech parameters relevant to motor disordered speech

Voice Range Profile (VRP)™, Model 4326, generates a phonetogram, or 2-dimensional plot of amplitude by fundamental frequency.

**Source:** Tiger Electronics, Inc. P. O. Box 85126, Seattle, WA:

Dr. Speech Science™ is a PC-based speech/voice assessment and training system for IBM compatible computers. It includes programs for speech analysis, speech training, voice assessment, voice synthesis and therapy, and EGG assessment.