An acoustic analysis of the effects of surgical therapy on voice quality

GERALD S. BERKE, MD (by invitation), **BRUCE R. GERRATT, PhD** (by invitation), and **DAVID G. HANSON, MD,**
Los Angeles, California

An automated computer system was used to measure changes in vocal quality resulting from laryngeal surgery. Five acoustic parameters were computed and seven listeners made judgments on vowel samples recorded before and after surgery from 37 patients. Presurgery and postsurgery changes in the five acoustic measures were compared with the perception of change in vocal harshness. A multiple regression analysis showed that four of the five acoustic measures accounted for a large amount of the variance in the judgment of change of voice quality following surgery. The measures may provide a quick and relatively sensitive method that may be clinically useful in measuring change in vocal quality. (OTOLARYNGOL HEAD NECK SURG **91**:502, 1983.)

In the last two decades, considerable progress has been made in phonosurgical techniques for the restoration, or enhancement, of voice quality. These include advances in microsurgery, implantation and injection of alloplastic material, the use of lasers, surgery of the laryngeal cartilages, and neuromuscular manipulation of laryngeal function.[1] The laryngologist's expanded capability to effect subtle changes in vocal function has led to new interest in vocal rehabilitation and assessment.

Consequently, the need for objective and reliable means of measuring changes in vocal quality is more apparent than ever to both the laryngologist and speech pathologist. While much can be learned about vocal pathologies by the perceptual evaluation of vocal quality, perceptual judgments are notoriously unreliable and difficult to quantify. Spectrographic analysis has been used in clinical settings for documentation of vocal characteristics, but it is also somewhat subjective and provides little data concerning actual laryngeal function.[2] Other forms of acoustic analysis have been proposed for a number of years but have been understood and evaluated by only a relatively small number of researchers.

As access to computer facilities increases, the application of acoustic signal processing techniques in patients with voice problems has become feasible in a clinical setting. In recent years a number of authors have used computers to digitize and analyze phonatory signals to identify the acoustic characteristics of normal and pathologic vocal function. Several acoustic features have been proposed as potentially useful in the characterization of vocal abnormality. These include measurements of pitch perturbation, amplitude perturbation, and spectral noise level.[2]

Pitch perturbation, sometimes referred to as jitter, is a measure of the degree to which the fundamental frequency varies from period to period. A lesion of the vocal cord may reduce the periodicity of vibration, an alteration that is correlated with listener perception of harshness or roughness.[3,4] Perturbation in the amplitude of period to period glottal vibration, referred to as shimmer, has also been related to the perception of vocal roughness.[4] Measurement of the amount of energy present as noise relative to the energy of the fundamental frequency and its harmonics (spectral noise level) and other measures of signal-to-noise ratio also have appeared promising as indicators of vocal pathology.[5,6]

From the UCLA School of Medicine and the West Los Angeles VA Medical Center.

Presented at the Annual Meeting of the American Academy of Otolaryngology—Head and Neck Surgery (Research Forum), New Orleans, La., Oct. 17-21, 1982.

Submitted for publication Oct. 17, 1982; accepted Feb. 15, 1983.

Reprint requests: David G. Hanson, MD, Section of Head and Neck Surgery, Los Angeles VA Medical Center, Los Angeles, CA 90073.

Major advances in automated computer signal processing have allowed vocal pathology to be studied in great detail at a reasonable expense. Linear predictive coding (LPC) is a widely used speech signal analysis method that is based on a model of the acoustic production of speech proposed originally by Fant.[7] A detailed discussion of the acoustic model of speech production is beyond the scope of this article; however, some orientation may be helpful.

According to Fant,[7] a vowel sound has the properties of the voice source modified by the filtering action of the vocal tract. The autocorrelation method of LPC analysis calculates the present speech sample as a linear combination of a number of previous samples. A set of predictor coefficients is used to approximate this vocal tract filter. By using an inverse of this filter, one can subtract the filtering effects of the vocal tract from the speech waveform and thereby estimate the voice source for a vowel sound.[8]

This derived excitation source, called the residue signal by Markel,[9] has been shown to be useful in the analysis of pathologic voices. Koike and Markel[10] reported that for some early and intermediate cases of laryngeal pathology the residue signal showed qualitative evidence of pathology, whereas the speech signal that had not been analyzed by LPC could not be differentiated from normal.

Davis[11,12] used a number of measures of the residue signal produced by LPC analysis to detect laryngeal pathology from recordings of vowels produced by speakers. He found these measures to be useful in distinguishing normal from pathologic voices. Since the measures were shown to identify relatively subtle degrees of vocal pathology, it was of interest to determine if they are sensitive enough to follow the effects of vocal therapy. Consequently, the present study was designed to assess the effectiveness of some of the measures employed by Davis in measuring change in vocal quality resulting from laryngeal surgery.

METHODS

Thirty-seven patients having vocal cord pathologies took part in this study. Their ages ranged from 24 to 72 years (mean 50 years). Eight patients had vocal nodules, 21 patients had vocal polyps, and eight patients had neoplasms confined to the true vocal cords. Treatment included cord stripping, laser excision, or radiation therapy. Many of the patients also received voice therapy from a speech pathologist. Twenty of the subjects were voice patients who presented to the UCLA School of Medicine, Division of Head Neck Surgery (UCLA and VA Medical Center, West Los Angeles), whereas 17 voices were studied from the group of patients that comprised source III in Davis' study.[11] In addition, the voices of 25 normal subjects without vocal pathology were recorded at UCLA. The normal subjects had an age range of 21 to 61 years and a mean of 39 years of age.

All subjects' voices recorded at UCLA were instructed to sustain the vowel /ɑ/ for as long and steadily as possible. Subjects were asked to produce three of these vowels, which were recorded on a high-quality audio recording system. All subjects' voices in the pathologic group were recorded once within 2 days of medical treatment and were recorded again within a 6-to 10-week period following treatment.

A 2-second segment from the middle portion of the longest vowel was selected from each subject's productions. This segment was low-pass filtered at 4.8 kHz and digitized at 10,000 times/sec with the use of a 12-bit A/D converter interfaced with an LSI 11/23 computer.

Analysis of the digitized speech segments utilized the computer algorithm described by Davis.[11] Essentially, this system uses an LPC method, reported in detail by Markel and Gray,[8] to compute the residue signal from which a number of acoustic measures were calculated. First, a Hamming window was applied without preemphasis filtering. Fourteen linear prediction coefficients were then calculated for male subjects and 10 coefficients were calculated for female subjects. These two filter orders were chosen empirically to provide clearly defined formant peaks in the linear predictive spectra. The length of the filter was 25.6 msec, while the acoustic measures were calculated over a 100 msec window of the speech signal. To minimize the influence of momentary changes in vocal quality, each digitized speech segment was analyzed over 10 consecutive 100 msec windows and the values obtained were then averaged. This procedure yielded mean values for the acoustic measures which were stored for further analysis.

Acoustic measures. Five acoustic measures were computed. Two perturbation measures were calculated, yielding a pitch perturbation quotient (PPQ) and an amplitude perturbation quotient (APQ). PPQ is analogous to the relative average perturbation reported by Koike[13]

$$\mathrm{RAP} = \frac{\dfrac{1}{N-2}\displaystyle\sum_{i=2}^{N-1}\left|\dfrac{P(i-1)+P(i)+P(i+1)}{3} - P(i)\right|}{\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} P(i)}$$

in which P(i), i = 1,2 . . . *n* over the successive pitch periods. The only difference was that this study used a

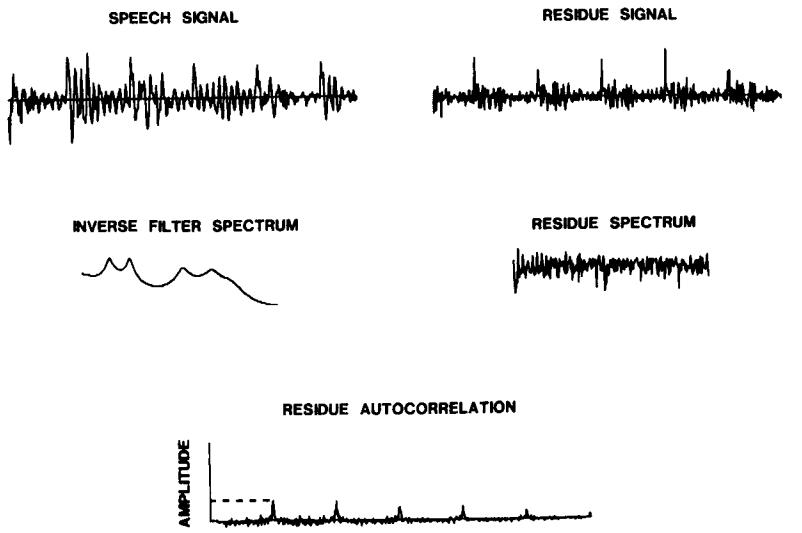

Fig. 1. Acoustic characteristics of synthesized vowel having equal amount of noise as voice in excitation source.

Table 1. Correlation of three acoustic measures to 1 dB increments in noise level of 15 synthesized vowels

Acoustic measure	Pearson r
SFF	.99*
SFR	.96*
PA	.97*

*$P < .001$.

five-point averaging window in the numerator instead of the three-point window used by Koike.[13] The APQ is analogous to the PPQ, except that amplitude values of each glottal cycle were used for each pitch period.

Two measures of spectral flatness were calculated: spectral flatness of the inverse filter (SFF) and spectral flatness of the residue signal (SFR). Spectral flatness is defined as the ratio in decibels of the geometric mean of the spectrum to the arithmetic mean of the spectrum.[11] The more the spectrum resembles noise, the greater its spectral flatness. The maximum value of spectral flatness measures for a signal composed entirely of noise is 0 dB. Thus, the more pathologic the voice, the more the amount of noise in the signal, and the closer the flatness measures will approach 0 dB. Davis[11] described the SFF as a measure of masking of the formants of the vocal tract by noise and the SFR as a measure of masking of the voicing harmonics by noise.

Another measure called the pitch amplitude (PA) by Davis[11] is assumed to represent the amount of voicing. He defined the PA as the value of the residual signal autocorrelation at the first pitch period peak. This is a measure of the degree of superposition of the residue signal upon itself after being shifted by one pitch period. Thus the less periodic and more noisy the signal, the less autocorrelation and the smaller the amplitude of the PA.

Validity of measurement. Three of the five acoustic measures which can be influenced by the amount of noise in the signal (i.e., SFF, SFR, and PA) were tested for their sensitivity to change in noise relative to signal (voicing). Fifteen /a/ vowels were synthesized by means of a computer program similar to that reported by Klatt.[14] All vowels had the same acoustic parameter values with the exception of the amplitude of frication (noise) in the sound source. The vowels were produced having 15 levels of noise differing in 1 dB increments. These 15 vowels were then analyzed in a manner identical to that reported for the subjects' vowels. The change in SFF, SFR, and PA was compared to the change in noise across the 15 vowels, with the use of Pearson's r. The higher the correlation, the greater the accuracy of these measures in assessing changing levels of noise in the vowels.

Perceptual analysis. To compare the results of the acoustic analysis with listener perception of vocal pathology, seven trained judges evaluated the normal and pathologic voice samples. The 99 digitized speech segments used in the acoustic analysis were converted to analog signals and recorded in a randomized order: 37 voices before surgery, 37 voices after surgery, and 25 normal voices. Four otolaryngologists and three speech pathologists first took part in a training session in which they were given instruction and judged normal

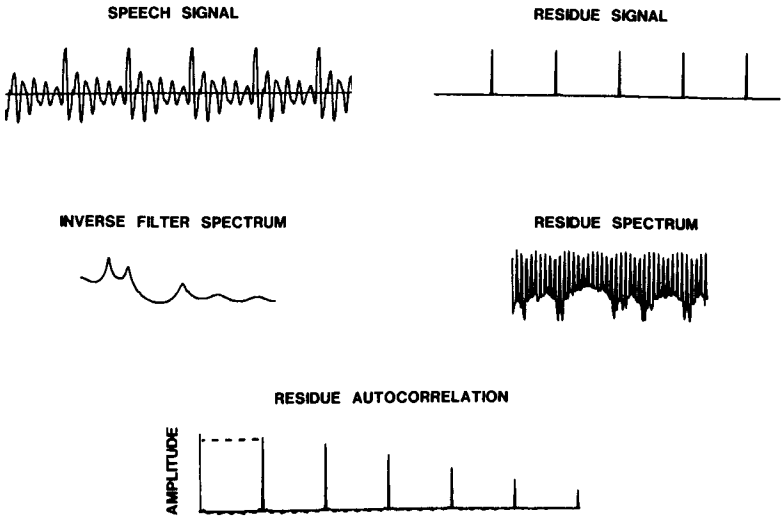

Fig. 2. Acoustic characteristics of synthesized vowel having all voice and no added noise in excitation source.

and pathologic synthetic vowels as well as vowels produced by patients not included in this study.

Following the training session, the judges were presented each vowel segment, preceded by a 5-second pause, three times in succession. To assess intrajudge reliability, the vowels of 10 subjects were randomly selected and presented again for evaluation. Judges were asked to score on a five-point, equal-interval scale of severity for vocal harshness, with one being normal and five most severely deviant. The vowel productions were presented in a sound field, in a sound-treated room at a comfortable listening level.

RESULTS

Fig. 1 demonstrates the effect of noise in the excitation source in one of the synthesized vowels. The residue signal is composed of both voicing impulses and much random energy occurring between these impulses. In contrast, Fig. 2 shows a synthesized vowel having no added noise in the excitation source. The residue signal of this vowel is composed entirely of evenly spaced spike trains. The use of the inverse filter has successfully removed the damped oscillations in the speech signal representing the filter effects of the vocal tract, leaving only the impulses of the voice source. A comparison of the spectral envelopes between the two vowels for both the inverse filter and the residue signal reveal differences. The two spectra of the noiseless vowel have more clearly defined peaks in their envelopes than the relatively flat envelopes of the noisy vowel. Another difference can be seen in the size of the first peak in the residue autocorrelation signal. The autocorrelation peak for the noiseless vowel is much larger than that for the noisy vowel.

Table 2. Comparison of 37 patients before laryngeal surgery and 25 normal subjects

Measurement	Subject category	Mean	SD	t	P value
PA	Normal	0.69	0.12	5.10	.000
	Before therapy	0.49	0.18		
PPQ(%)	Normal	2.50	3.40	−2.33	.023
	Before therapy	4.77	4.07		
APQ(%)	Normal	10.00	2.75	−3.42	.001
	Before therapy	12.54	2.98		
SFR(dB)	Normal	−10.10	2.60	−4.16	.001
	Before therapy	−7.17	1.41		
SFF(dB)	Normal	−17.13	2.83	−1.14	.258
	Before therapy	−16.71	4.21		

Measurement of these observations for the 15 vowels are presented in Table 1. These high correlations demonstrate that incremental change in the amount of noise in the excitation source of synthesized vowels is directly proportional to change in the values of PA, SFR, and SFF. Such a relationship implies a high degree of validity for these three parameters in measuring degree of noise.

The interjudge reliability with the use of the intraclass correlation was 0.86. Thus there was moderately high agreement among the judges in their per-

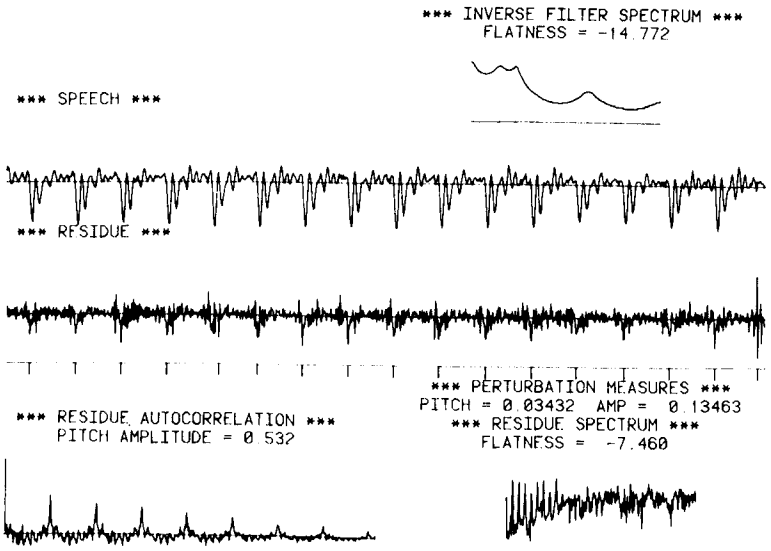

Fig. 3. Acoustic characteristics of vowel produced by female patient before therapy.

Table 3. Correlation of five acoustic measures to perceptual judgment of vocal harshness

Acoustic measure	Pearson r
PA	.80*
PPQ	.73*
APQ	.66*
SFR	.61*
SFF	.23

*$P < .001$.

ception of vocal harshness. For intrajudge reliability, Pearson's r was used to compare the first set of judgments to the second set. A correlation of 0.87 was found, demonstrating that the raters' judgments were stable over time.

Table 2 shows *t*-test comparisons of the normal and pathologic groups before surgery. Four of the five acoustic measures were able to distinguish between normal and pathologic voices. These results are similar to those reported by Davis[11,12] that acoustic measures derived from the residue signal can differentiate between these two groups.

Beyond determining whether these measures can identify pathologic vocal production, we were interested in finding if they also provide a sensitive tool to assess the effects of therapy. As a first step, correlations were calculated between the perceptual judgments of the pathologic voices recorded before therapy and their respective acoustic measures. Table 3 shows that the same four measures that distinguished between the pathologic and normal groups are moderately related to the judgment of vocal harshness. Of these four measures, PA is the most strongly correlated, while SFF is most weakly correlated.

A more important consideration is the sensitivity of these acoustic measures in assessing change in vocal quality. Consequently, for each pathologic subject, the difference in the acoustic measures from presurgery to postsurgery was compared with the difference occurring in the perceptual judgments of vocal quality over this period. A multiple regression analysis produced a regression equation, in which the most important acoustic measure in predicting change in vocal harshness was PA, followed in order of importance by PPQ, SFR, and APQ. SFF was the least important.

Table 4 presents a correlation matrix of this difference data. Since there is a moderate degree of intercorrelation among the acoustic measures, the multiple R value of .86 probably has been inflated by suppression among these predictor variables and should be interpreted cautiously. The acoustic measures probably account for less than the 74% of the variance in the perceptual judgments indicated by the multiple R.

Both PA and PPQ are moderately related to change in the perceptual judgments. In fact, when taken individually, either PA or PPQ can account for approximately 50% of the variance in the perception of change of vocal harshness. However, the moderate intercorrelation of .74 between PA and PPQ implies that they are measuring similar phenomena.

Figs. 3 and 4 demonstrate results of the acoustic analysis before and after therapy for one female patient. She had been treated for vocal nodules by laser mi-

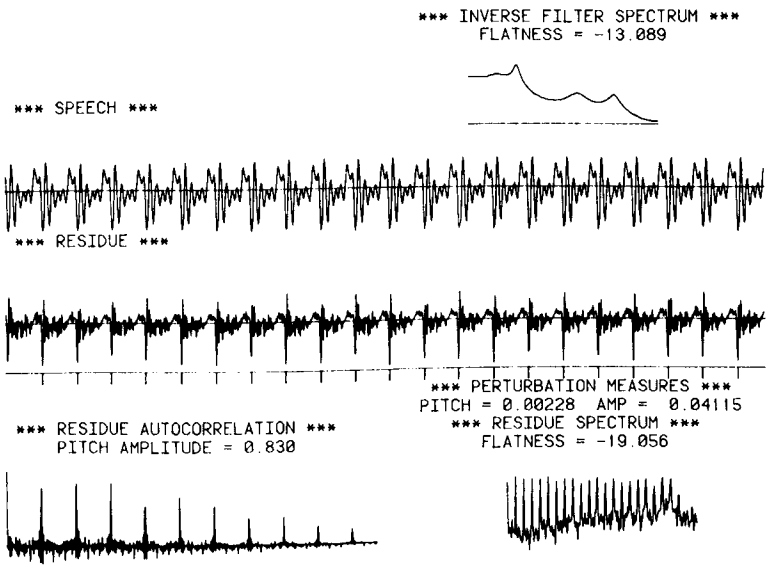

Fig. 4. Acoustic characteristics of vowel produced by female patient after therapy.

crolaryngoscopy and had received voice therapy by a speech pathologist. Notice the increased periodicity and reduced level of noise between voicing impulses in the residue signal after therapy. Comparison of the speech signal alone does not portray the differences that can be seen in the residue signal. Furthermore, both perturbation measures (PPQ and APQ), the peak in the autocorrelation (PA), and one of the flatness measures (SFR) have improved noticeably following therapy.

DISCUSSION

The analysis of voice characteristics by means of the autocorrelation method of LPC has several advantages. It is an automated procedure that produces objective and reliable measures of voice relatively quickly. In addition, data collection is noninvasive and quite simple to perform with a high-quality audio recording system.

PA appeared the most sensitive of the acoustic measures in the multiple regression analysis and was also most highly correlated with the perceptual judgments of harsh voice quality before therapy. PPQ also was found to be a fairly sensitive measure. These results are in contrast to those reported by Davis.[11] He found PA, APQ, and a signal-to-noise type measurement he called the coefficient of excess better able to differentiate the normal from pathologic groups. Although this study employed many of the same measures used by Davis, there were enough differences in methods between both studies to account for these different findings. One great difference was the use of mean values obtained from averaging over 10 consecutive windows. This procedure reduced the effects of extreme values occurring as a result of momentary fluctuations in the vocal tract. In addition, this study included only patients exhibiting vocal pathologies secondary to mass lesions of the vocal cords.

Table 4. Intercorrelations among change in perceptual and acoustic measures before surgery and after surgery

	Judgment	PA	PPQ	APQ	SFR	SFF
PA	−0.70*	—				
PPQ	0.71*	−0.73*	—			
APQ	0.38	−0.71*	0.69*	—		
SFR	0.48†	−0.88*	0.69*	0.74*	—	
SFF	0.14	−0.13	−0.07	0.01	0.14	—

*$P < .001$.
†$P < .01$.

On the other hand, SFF was shown in a number of analyses not to provide much information about voice. It did not differentiate between the normal and pathologic groups, it was weakly related to the judgment of harshness, and was the least important measure in the multiple regression analysis. Although SFF was found to be highly sensitive to small incremental change in noise added to the excitation source in synthesized vowels, it was ineffective in describing vocal characteristics of patients having mass lesions of the vocal folds. This may be explained by the fact that SFF is a measure based on vocal tract characteristics and therefore may not be highly correlated to this variety of vocal pathology. It is important to point out that pa-

tients having neuromuscular vocal dysfunction were not included in this study. It is quite possible that SFF may indeed provide important information regarding weak, breathy voices that may accompany neurogenic voice disorders.

Actually, a single acoustic measure cannot fully describe the many varieties of vocal pathologies, since there is not a clear relationship between any one acoustic measure and a particular pathology. For example, a high level of spectral noise may be caused in part by an abnormally large glottal gap during phonation or also may be caused by quite a large degree of aperiodic glottal vibration. Considering the complex physiology of vocal production, a multidimensional approach is helpful when using acoustic measures.

There are a variety of other acoustic measures from which to choose that also can provide information concerning laryngeal pathology. Four of the five measures used in this study have been shown to be related to the perception of change in voice quality and may be clinically useful in following the effects of therapy on vocal quality. However, their relationship to the physiology of vocal production is still unclear. Further research is needed to study the relationship of these measures to the actual behavior occurring at the glottis.

We wish to thank Steven B. Davis for the use of his computer programs and helpful suggestions. We thank June Shoup and the Speech Communication Research Laboratory for access to their audiotape library. We deeply appreciate the use of laboratory facilities and support provided by Douglas Noffsinger. The engineering and technical support of Steven M. J. Hunt and Robert Vanderveen is gratefully acknowledged. We also thank Peter Ladefoged for the use of his computer facilities and Gunnar Fant for his valuable advice.

REFERENCES

1. Isshiki N: Recent advances in phonosurgery. Folia Phoniatr **32:**119-154, 1980.
2. Childers CG: Laryngeal pathology detection. CRC Crit Rev Bioeng **2:**375-425, 1977.
3. Wendahl RH: Laryngeal analog synthesis of harsh voice quality. Folia Phoniatr **15:**241-250, 1963.
4. Wendahl RH: Laryngeal analog synthesis of jitter and shimmer auditory parameters of harshness. Folia Phoniatr **18:**98-108, 1966.
5. Lively M, Emanuel F: Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult females. J Speech Hear Res **13:**503-517, 1970.
6. Sansone R, Emanuel F: Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult males. J Speech Hear Res **13:**489-502, 1970.
7. Fant CGM: Acoustic theory of speech production. S'Gravenhage, 1960, Mouton & Co.
8. Markel JD, Gray AH: Linear prediction of speech. New York, 1976, Springer-Verlag.
9. Markel JD: Application of a digital inverse filter for automatic formant and Fo analysis. IEEE Trans Aud Electronics **21:**154-160, 1973.
10. Koike Y, Markel JD: Application of inverse filtering for detecting laryngeal pathology. Ann Otol Rhinol Laryngol **84:**117-124, 1975.
11. Davis SB: Computer evaluation of laryngeal pathology based on inverse filtering of speech. SCRL Monograph 13, Santa Barbara, 1976, Speech Communications Research Laboratory, Inc.
12. Davis SB: Acoustic characteristics of normal and pathologic voices. *in* Ludlow CL, O'Connell M, editors: Proceedings of the Conference on the Assessment of Vocal Pathology. Rockville, Md, 1981, ASHA Reports, pp 97-115.
13. Koike Y: Application of some acoustic measures for the evaluation of laryngeal dysfunction. Studia Phon **7:**17-23, 1973.
14. Klatt DH: Software for a cascade/parallel formant synthesizer. J Acoust Soc Am **67:**971-995, 1980.