Acoustic Analysis of Speech: Validation Studies

R. C. A. WEATHERLEY-WHITE, M.D.
RICHARD B. STARK, M.D.
CLAYTON R. DeHAAN, M.D.
New York, New York

For the past two years, the Plastic Surgery Service of St. Luke's Hospital has been engaged in a study of defective speech, using as subjects patients from the cleft palate clinic. One of the principal features of this study has been the clinical assessment of a new instrument designed by acoustic engineers to quantify degrees of nasality in connected speech. This instrument, the Voice Systems Nasality Meter, is based on a prototype speech recognition device of the IBM Corporation, and makes use of the phase, or time, relationships in the frequency spectrum of the speech wave to define degrees of nasal resonance. In a previous investigation (*13*), we studied the speech of 25 cleft palate patients and 25 pair-matched controls, obtaining a certain degree of correlation between the instrumental assessment with the Meter and a perceptual judgment derived from a listener panel.

We felt, however, that this correlation was not sufficient in itself to define the mode of action of the instrument, and that a more fundamental study, plotting the instrumental response to specific speech stimuli, was needed. By this means it was hoped to clarify what aspect of speech was recognized and measured by the Nasality Meter.

Method and Materials

The Voice Systems Nasality Meter has been described in previous publications (*12, 13*). Historically, it was derived from a speech recognition device which, as part of the logic for the transcription of a small vocabulary of spoken words, made use of the 'nasality' of vowels in a specific phonetic context (*3*). It was felt that this concept could be reversed to permit the assessment of nasality for a standard test phrase uttered by different speakers. Furthermore, as nasality has been described as one of the principal stigmata of cleft palate speech, it was hoped that the instrument would be of use in the classification of the speech performance of post-operative cleft palate patients.

Drs. Weatherley-White, Stark, and De Haan are affiliated with the Surgical Service (Plastic) of St. Luke's Hospital in New York City.

This paper was presented at the 1965 Convention of the American Cleft Palate Association in New York City, and was supported by Public Health Service Research Grant DE-02009, National Institute of Dental Research.

The speech signal is introduced into the instrument either live or on tape. The raw pressure wave is then integrated and subjected to phase-shift, producing an asymmetric wave-form via a low-voltage output. This may be observed by means of a meter which is part of the instrument, or may be recorded permanently by kymograph for which an output jack is provided. In the present study, the phase-shifted envelope was monitored on a Sanborn 569A oscilloscope, and when satisfactory adjustments had been made, recorded permanently by means of a Sanborn 322 direct-writer running at 25 mm per second. A second channel recorded the raw speech wave to correlate the envelope with vocal cord activity.

The speech sample subjected to integration and phase-shift was uttered by an adult white male without pathology of the speech producing mechanism and consisted of four disyllables chosen from a larger group of sounds used in the course of an electromyographic study being performed at the Haskins Laboratories (*6*).

These disyllables, *apple, napple, ample,* and *sample,* made use of a standard vowel /a/ positioned in varying phonetic contexts with particular reference to the effect of preceding and following nasal consonants. The /a/ was used rather than the more natural /æ/, because in the American dialect, /æ/ is subject to varying degrees of nasality. The instrumental response to these signals was recorded as described.

The temporal and spatial relationships of the speech producing mechanism during these utterances were recorded cineradiographically for the same subject using a 9-inch image intensifier and 16 mm film at 24 frames per second. Each frame for the entire sequence was traced on a frame-by-frame editor, care being taken to identify the anterior and posterior nasal spines, the body of the Atlas, the posterior pharyngeal wall, the outline of the soft palate, and the surface of the tongue.

For standardization, the baseline from anterior to posterior nasal spine was projected backward to a point on the posterior pharyngeal wall and measurements, after correction for parallax, of the velopharyngeal opening were taken from the point. The velopharyngeal gap, representing coupling of the oral and nasal resonating chambers, was plotted against time, as was the extent of velopharyngeal apposition. Observations were also made of tongue position and pharyngeal wall excursion.

Hence a graphic representation of velar movement was obtained; and if one can assume that the velopharyngeal gap measured from a lateral projection bears a relationship to the cross-sectional area of the open port, then the degree of coupling of the oral and nasal resonating chambers could be inferred for any moment in time of the speech events studied.

Results

The relationships of the tongue, soft palate, and posterior pharyngeal wall with the palate at rest are shown in Figure 1. These structures are

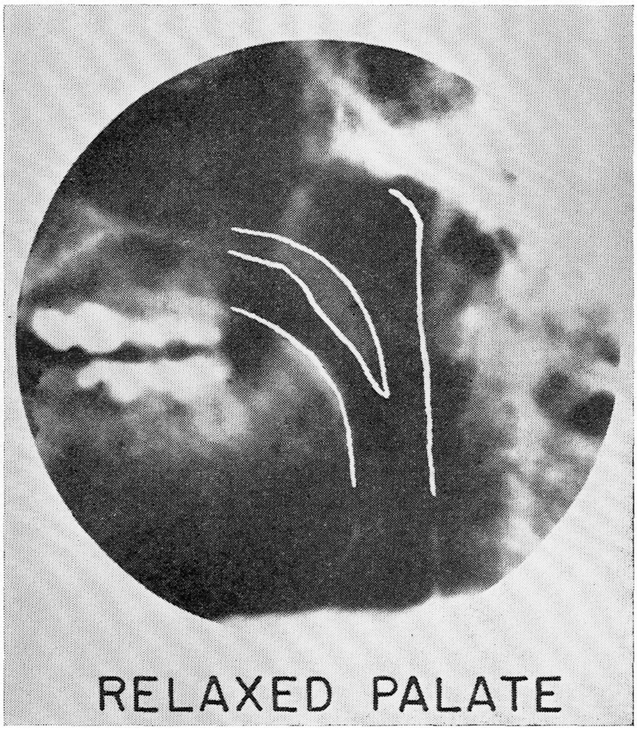

FIGURE 1. A single frame from the cineradiograph showing the structures at rest. The tongue, soft palate, and posterior pharyngeal wall are outlined.

outlined, as single frames from a film strip lack the precise definition of still X-rays. In this case, the velopharyngeal aperture, measured from the backward projection of the hard palate to a point on the posterior pharyngeal wall, is 12 mm with the palate at rest.

CONDITION 1 (APPLE). The utterance of *apple* is initiated by a posterior and upward movement of the tongue. About 400 milliseconds prior to voicing, the palate moves briskly to contact the posterior pharyngeal wall just above the level of the Atlas. At the precise moment of contact, voicing of /a/ begins (Figure 2). The onset of speech is delineated on the graph by the vertical arrow, and the relationships of the speech producing structures at that moment are portrayed in the inset photograph.

To achieve the degree of closure necessary for plosive production, a second more forceful contracture of the levator muscles occurs. Assisted by superior constrictor activity, the palate is opposed to the posterior pharyngeal wall for about 15 mm of its length as the /p/ is uttered. As vocal cord activity ends, both palate and posterior pharyngeal wall return to the resting position until the sequence is repeated.

Figure 3 shows the kymographic tracing of the integrated and phase-shifted speech wave correlated with palatal movement as extrapolated

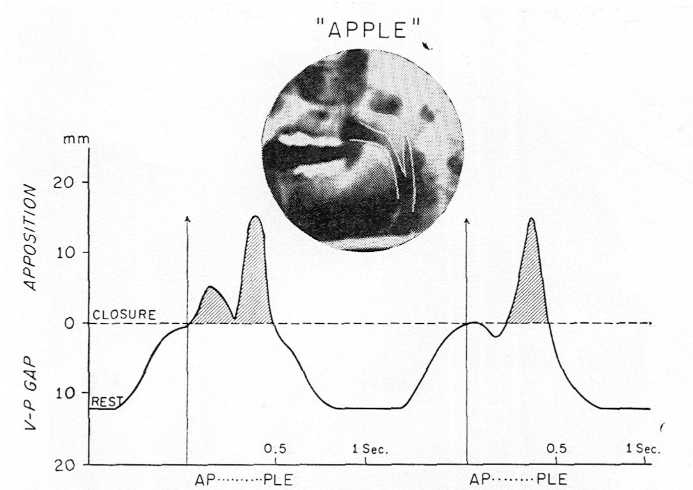

FIGURE 2. A graphic representation of velar movement during production of the di-syllable word *apple*. Both the velopharyngeal gap and the extent of velopharyngeal closure are plotted continuously against time. Inset is a cine frame showing palatal position at the initiation of sound.

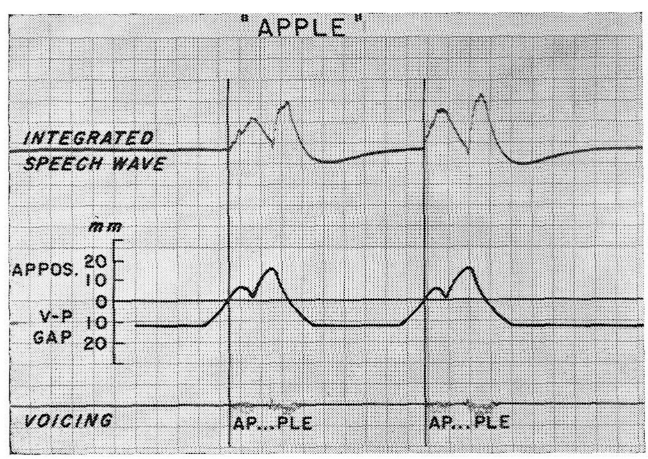

FIGURE 3. The tracing of palatal movement is superimposed on the kymograph of the integrated phase-shifted speech wave. The pressure wave is seen at the bottom on a second channel.

from the previous graph. The envelope produced by the Nasality Meter responds with a biphasic wave corresponding to the two syllables of *apple*, both peaks evidencing positive polarity.

Condition 2 (Napple). In this condition, a nasal consonant initiates voicing, and the palate is not approximated to the pharyngeal wall at the onset of speech (Figure 4). There can be seen, both on the tracing and in the inset photograph, about 8 mm of antero-posterior coupling of the oral and nasopharynx which produces nasal resonance for the /n/

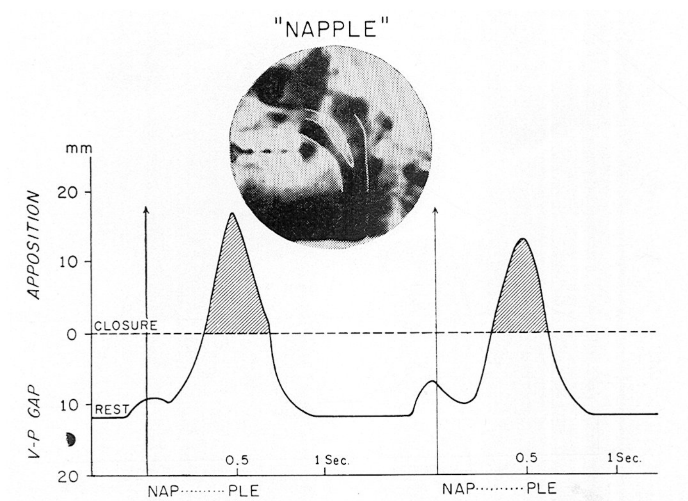

FIGURE 4. Palatal movement during production of the di-syllable *napple,* with cine frame showing velar position at the onset of speech.

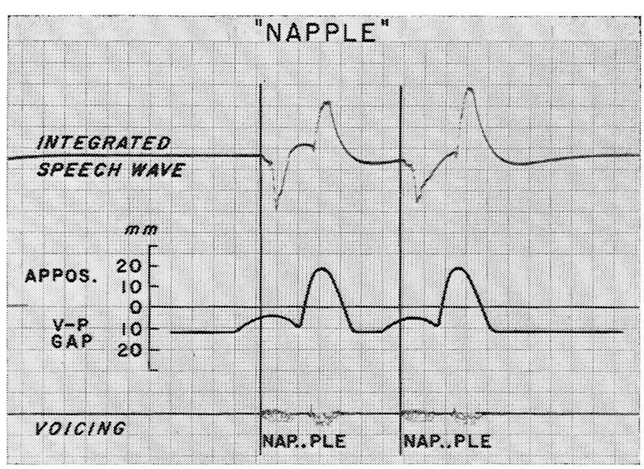

FIGURE 5. The integrated speech wave shows an initial negative polarity for the word *napple,* corresponding to oronasal coupling defined from the tracing of palatal movement.

sound. With the transition to the orally produced second syllable, palatal activity is quite similar to the preceding sequence of events.

In Figure 5, the phase-shifted envelope shows initial negative polarity, but as the nasal resonating chamber is excluded from the system by the rapidly contracting palate, the wave-form becomes positive as before, the two modalities coinciding quite precisely.

CONDITION 3 (AMPLE). In this condition, the nasal consonant /m/ follows the initial /a/ (Figure 6). The palate rises to barely oppose the

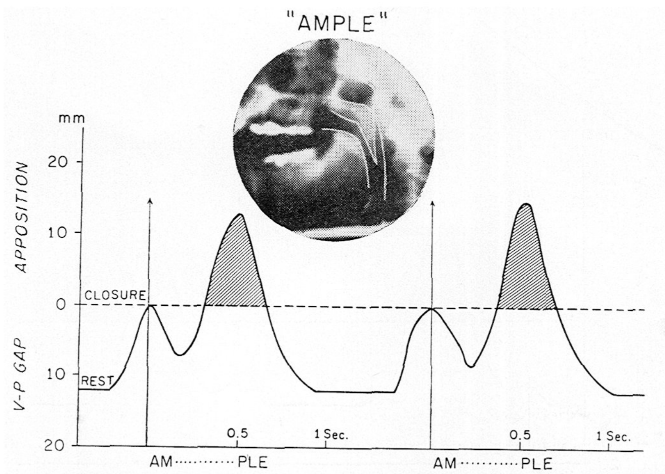

FIGURE 6. Palatal movement during the sound *ample* demonstrates a close-open-close sequence at the velopharyngeal port.

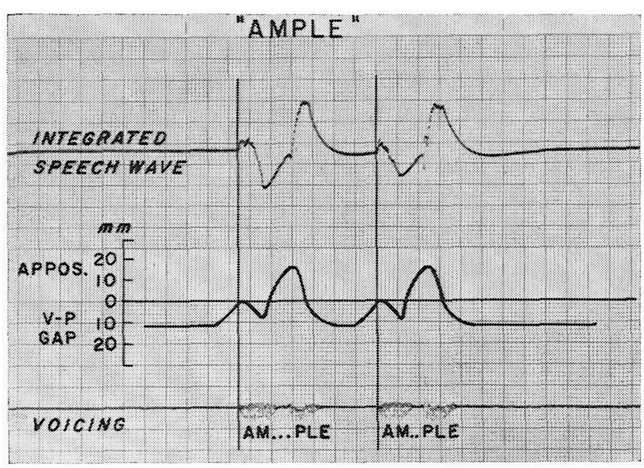

FIGURE 7. The polarity of the phase-shifted speech wave mirrors the relationships of the oral and nasal resonating chambers.

posterior pharyngeal wall at the onset of voicing, then relaxes to allow nasal resonance to occur on the /m/ before its burst of plosive activity.

The wave form produced by the Nasality Meter faithfully reproduces the changes occurring in the resonating system, being positive-negative-positive as the relationships alter (Figure 7).

Condition 4 (Sample). Production of an initial sibilant requires a greater degree of velopharyngeal apposition than does an oral vowel. The palate and superior constrictor working synchronously are seen to achieve about 2 cm of apposition before voicing takes place (Figure 8). The transition from nasal consonant /m/ to plosive /p/ requires a rapid

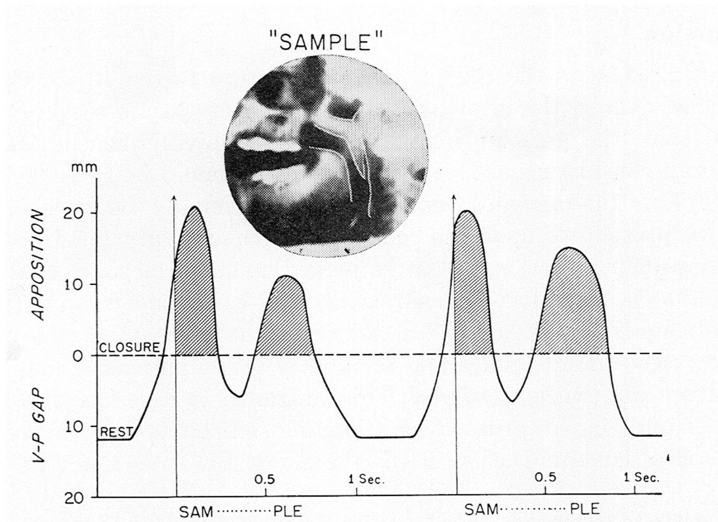

FIGURE 8. Both the inset cine frame and the tracing of palatal movement show the great extent of velopharyngeal closure required for sibilant production.

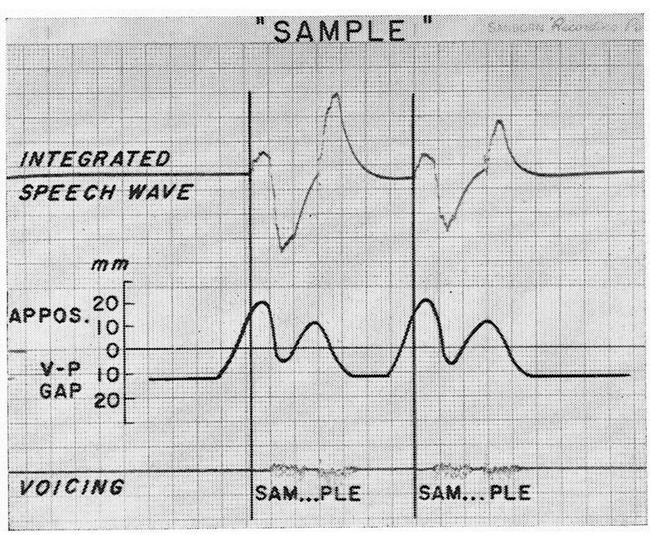

FIGURE 9. The polarity of the phase-shifted speech wave again is determined by the degree of coupling of the oral and nasal resonating chambers.

relaxation and contracture of the palate. This 'close-open-close' sequence takes place in about 300 milliseconds. The time factor of this capability appears quite standard for normal speakers and would appear to be a useful criterion for defining normal palatal mobility.

The phase-shifted envelope in Figure 9 shows an initial irregular low positive peak representing sibilant production preceding vocal cord activity, and then precisely mirrors the rapid sequence of coupling and uncoupling of the oral and nasopharynx as defined by the cine study.

Discussion

Wardill (*11*), in the 1928 Hunterian Lecture to the Royal College of Surgeons, defined the priorities involved in cleft palate surgery when he stated that 'the greatest item . . . must undoubtedly be the speech results, and considerations . . . largely aesthetic should be given a very low value'. This fundamental concept must continually be borne in mind. Surgical procedures upon the congenital cleft are truly functional in nature, and their success must not be judged upon the elegance of the repair or even upon physiologic measurements that can be made using the many sophisticated devices which are today at the disposal of the clinician. Rather, they must be judged on the ability of the patient to communicate effectively and without stigma. This judgment is necessarily subjective, and related to perceptual evaluation, for human speech, being a phenomenon of communication, needs the ear of the listener for its recognition.

In order to compare the results of different centres, some reproducible criteria of speech proficiency must be evolved. The disparate disciplines involved in the care of the cleft palate patient have all contributed to our understanding of this problem.

The fields of speech pathology and experimental psychology have given us insight into rating techniques permitting standardization of perceptual judgments, which, when well controlled, give good statistical correlation between different panels of judges (*9*). Pressure (*2*) and flow (*10*) studies have defined the competence of the velopharyngeal valving mechanism, the defectiveness of which lies at the very heart of unintelligible cleft palate speech.

Cineradiography portrays the moving parts of the speech-producing mechanism (*7*), and with frame-by-frame tracing techniques, permits measurements to be made of the relationships of these structures during speech.

Whereas measurements of pressure phenomena and anatomic relationships provide indirect information about speech production, the physical properties of speech itself may be quantitated acoustically. The most familiar method of speech analysis on the acoustic level is by means of the sound spectrograph, which displays the power components of speech in a visible pattern, plotting both frequency and intensity against time. Cross-sectional 'cuts' of the tracing can be made, enabling the relative intensity of different frequencies at a specific instant to be quantitated.

The most definitive work to date on the application of this technology to cleft palate speech has been performed by Nylen (*8*) and Bjork (*1*) in association with Gunnar Fant. They have shown, by means of spectrography synchronized with cineradiography, that articulatory proficiency is directly related to velopharyngeal competence. As the cross-sectional area of the palatopharyngeal port increases, the intensity of high-frequency bursts (principally the plosives, fricatives, and affricates) is attenuated. Their studies relating nasality to formant intensity and bandwidth were less productive, the immense variability of normal individuals precluding any true standardization of the spectrographic response to voice quality defectiveness. This finding has been confirmed by other investigators (*4, 5*).

The voice quality deficit in severely handicapped cleft palate speech is more complex than implied by the term 'hypernasality'. Breathiness, huskiness, throatiness, and flatness all contribute to the characteristic spectrum of impairment of these speakers. The inability of the incompetent velopharyngeal valving mechanism to exclude the nasal resonating chamber from the speech-producing mechanism with a consequent overlay of excessive nasal resonance plays the most prominent role, however. Observation of the wave-form of the integrated, phase-shifted speech envelope in this investigation demonstrates a quite distinct and consistent response to nasal resonance. When the cineradiograph indicates coupling of the oral and nasal resonating chambers by means of the open velopharyngeal port, the polarity is consistently negative.

When the nasopharynx is excluded from the system by palatopharyngeal apposition, the polarity is positive. Rapid changes in resonance are faithfully recorded. It would appear from the foregoing study that the precise response to changes in degrees of nasal resonance achieved by the phase-shifting circuitry of the Nasality Meter would justify its continued use in further investigations. Direct acoustic measurement of voice quality will be useful in studies of normal speech production, in linguistic studies of variations in normal speech patterns, and when correlated adequately with properly supervised perceptual evaluations, in the assessment of defective speech.

Summary

The acoustic recognition of faults in articulation by the spectrograph is more reliable than that of defects in voice quality. An attempt has been made to study the response of the Voice Systems Nasality Meter to changes in nasal resonance with sounds of known acoustic characteristics. The integrated, phase-shifted envelope of *apple, napple, ample,* and *sample* was plotted against coupling of the oral and nasal resonating chambers as derived cineradiographically. A distinct and consistent response to changes in nasal resonance was obtained, the wave-form showing positive polarity when the nasopharynx was excluded by palatal apposition, and negative polarity when nasal resonance was achieved via the open velopharyngeal port. The ability to thus measure nasal resonance is deemed useful in acoustic studies of the cleft palate speech.

reprints: *R. C. A. Weatherley-White, M.D.*
St. Luke's Hospital
Amsterdam Avenue and 113th Street
New York 25, New York

Acknowledgment: The authors would like to thank Katherine S. Harris, Ph.D., Professor of Speech Science at Hunter College and Research Associate in Experimental Psychology at the Haskins Laboratories, New York City, whose wise counsel has catalyzed a fruitful and stimulating association.

References

1. Björk, L., Velopharyngeal function in connected speech: Studies with tomography and cineradiography synchronized with speech spectrography. *Acta Radiol.,* Suppl. *202,* 1–94, 1961.
2. Chase, R. A., An objective evaluation of palato-pharyngeal competence. *Plastic reconstr. Surg., 26,* 23–29, 1960.
3. Dersch, W. C., A decision logic for speech recognition. *IBM Technical Report:* 16-01-106-018: 1961.
4. Dickson, D. R., An acoustic study of nasality. *J. speech hearing Res., 5,* 103–111, 1962.
5. Hanson, M. L., A study of velopharyngeal competence in children with repaired cleft palates. *Cleft Palate J., 1,* 217–231, 1964.
6. Harris, Katherine S., Rosov, R., Cooper, F. S., and Lysaught, G. F., A multiple suction electrode system. *Electroenceph. clin. Neurophysiol., 17,* 698–700, 1964.
7. Moll, K. L., Cinefluorographic techniques in speech research. *J. speech hearing Res., 3,* 227–241, 1960.
8. Nylen, B. O., Cleft palate and speech: A surgical study including observations on velopharyngeal closure during connected speech, with synchronized cineradiography and sound spectrography. *Acta Radiol.,* Suppl. *203,* 1–124, 1961.
9. Prather, Elizabeth M., Scaling defectiveness of articulation by direct magnitude-estimation. *J. speech hearing Res., 3,* 380–392, 1960.
10. Quigley, L. F., Jr., Shiere, F. R., Webster, R. C., and Cobb, C. M., Measuring palatopharyngeal competence with the nasal anemometer. *Cleft Palate J., 1,* 304–313, 1964.
11. Wardill, W. E. M., Cleft palate. *Brit. J. Surg., 16,* 127–148, 1928.
12. Weatherley-White, R. C. A., Dersch, W. C., and Anderson, Ruth M., Objective measurement of nasality in cleft palate patients: A preliminary report. *Cleft Palate J., 1,* 120–124, 1964.
13. Weatherley-White, R. C. A., Stark, R. B., and De Haan, C. R., Objective measurement of nasality in cleft palate patients: Correlation with listener judgments. *Plastic reconstr. Surg., 35,* 588–598, 1965.