Physio-Acoustic Considerations in the Radiographic Study of Speech

JOANNE D. SUBTELNY, Ph.D.
Rochester, New York

Definitive discussion of roentgenography applied to the study of palatopharyngeal function should be prefaced by a statement of the purpose of that study and the parameters selected for investigation. If the parameters selected require a specification of the size and shape of the pharynx and a specification of oral and pharyngeal soft tissues, including lymphatic tissue, then cephalometric roentgenography is recommended.

Cephalometric Roentgenography

In clinical practice, a primary objective is to quantify the degree of palatopharyngeal opening, the degree or level of velar elevation, and the vertical dimension over which palatopharyngeal contact is established. For these needs, cephalometric roentgenography (still films) is the procedure of choice. Subtle differences in velopharyngeal relationships or physiologic details of small magnitude are revealed best by the cephalometric headplate.

Some features of cephalometric roentgenography which recommend its use for the physiologic parameters designated here will be enumerated. Cephalometric roentgenography provides a rigorous standardization of technique making it possible to control magnification and distortion, both of which are inevitable in radiographic films. There is no appreciable movement occurring during exposure. As a result, soft tissue relationships are delineated clearly. These features make quantitative measurements in the palatopharyngeal region possible and applicable to a well-defined skeletal reference. In summary, cephalometric roentgenography is a reliable, valid, relatively accessible, and economical technique. Its use, in conjunction with other diagnostic procedures, is highly recommended for assessment of palatopharyngeal competence.

Although the research value of cephalometric films possibly has been exploited in study of palatopharyngeal function, the technique may be utilized to study other areas of interest to speech physiologists. Sound

---

This paper was presented as part of a symposium on radiological tools in velopharyngeal studies at the 1963 Convention of the American Cleft Palate Association, Washington, D. C.

Dr. Subtelny is Research Associate, Orthodontic Department, Eastman Dental Dispensary, Rochester, New York. The research was supported in part by research grant DE-01837, National Institutes of Health, United States Department of Public Health.

recordings, synchronized with x-ray exposure, may be secured to control the phonetic accuracy and stability of the utterance. Static references derived are informative and assist in analysis and interpretation of cineradiographic films.

The most obvious limitation of the cephalometric film is its restriction to a static appraisal of function. Static films, revealing velopharyngeal relationships during sustained sound production, have predictive value in defining potential for achieving closure during continuous speech. However, the study of continuous speech function requires cineradiography.

Cineradiography

With the cine technique, a measurable graph of function can be obtained and thus, diagnosis of palatopharyngeal function may be improved materially.

In regard to the total activity occurring within the vocal tract during speech, there is almost an overabundance of information registered on the cine film. Patterns of activity have been described for some speech structures during normal articulation of prescribed phonetic sequences. However, many physiologic details of articulation and their respective significance to speech output have not been defined by organized research. For this reason, continued study is needed before confident interpretations of observed activity can be made. Prerequisite to identification and interpretation of abnormal speech function, variability in normal speech activity merits careful investigation.

Variability in Articulatory Behavior

There are at least two major sources of variability in normal articulatory behavior which merit critical study: a) variability incident to modification in phonetic context, and b) variability associated with structural deviations.

Phonetic Context. Physiologic and acoustic features of a specific sound are modified to some extent by the phonetic environment of the sound and the physiologic factors associated with the production of sounds in sequence. In other words, the direction and amount of physiologic movements required to produce a given sound are influenced by physiologic parameters of preceding and succeeding sounds. Spectral modifications resulting from these movements have been studied extensively from the acoustical viewpoint. Further research has shown that the spectral modifications provide auditory cues which are highly important to perception.

The extent of physiologic change which can occur as a function of phonetic context is well illustrated by recent research. High speed motion pictures and spectrographic study have shown that when the /r/ sound is preceded by a voiceless plosive, as in the syllable /pro/, the /r/

sound is produced without laryngeal vibration (*4*). The /r/ sound in most phonetic contexts is produced with voicing or laryngeal vibration. This isolated research report is included to demonstrate modifications in physiologic determinants of a sound which can occur as a function of phonetic context. Such modifications may offer a partial explanation for clinical observations that cleft palate speakers can produce a particular sound accurately in one sequence or word and yet be unable to produce the same sound in another word or different phonetic context.

Physiological modifications, determined by phonetic environment, should be acknowledged in research planning and investigated. The study of movements and the patterning of movements within controlled phonetic contexts seems a logical focal point for cine research.

Deviations in Structure. Variability in function also might be anticipated as a consequence of structural differences. Modifications in function have been described when gross deviations in form, such as cleft palate, exist. Lesser degrees of structural variations such as differences in the skeleto-dental environment and mouth shape, might also be expected to exert some influence upon the functional patterns observed. Some modifications in function also may be attributed to differences in status of growth and development. For example, the site of palatopharyngeal closure and the height of velar elevation are known to change as a function of age. For these reasons, articulatory movements should be studied within a carefully specified structural reference.

Perhaps it should be mentioned that variant articulatory patterns will not have clinical significance unless the observed patterns are found to be associated with defective speech production. For this reason, normal and defective patterns of movement should be differentiated by evaluative judgments of synchronous speech recordings. In essence, differences in anatomic and physiologic details (form and function) must be adjudicated within the reference of speech adequacy before such differences can be considered to make a difference.

More research similar to that contributed by Powers (*18*) is needed to identify and differentiate normal and abnormal speech coordination when normal and abnormal morphologic features are present. In short, continued research is needed before the full potential of the cine film for speech diagnosis can be realized. To illustrate, diagnostic value of cephalometric films increased as a function of research data supplied by their use. A wealth of information was recorded consistently; however, certain details, such as cranial base angulation, did not become meaningful in diagnosis until research established their significance.

The Emergence of Cine Diagnostic Reference

The development of a diagnostic reference for the cine film might well be slow in emergence for several reasons. The technique is complex and comparatively new; it has been and still is, going through technological

refinements. Furthermore, cineradiographic studies of speech have not demonstrated standardization of equipment, procedure, and/or method of analysis. Although this fact is completely comprehensible, it should be acknowledged as a detriment to the development of a normative speech reference.

Research experience of cineradiographers studying deglutition has shown that the variable of frame rate, by itself, can have an appreciable influence upon the interpretation of films. Cineradiographic studies of speech indicate that frequency of frames per second is just one feature in the procedure which has been varied.

Differences in phonetic structure of the speech samples studied are additional variables. As previously mentioned, phonetic details will influence the patterns of movements observed, during continuous speech. The *duration* of specific sounds is also modified by phonetic context (*8*). These facts again emphasize the necessity of acknowledging phonetic principles in research planning to standardize cine procedure.

The importance of controlling phonetic structure is illustrated by recalling the fact that each vowel is produced by rather consistent and characteristic tongue, velar, jaw, and lip positions. During normal production of low (open) vowels, the velum is comparatively low and may not necessarily contact pharyngeal tissue. High (closed) vowels are produced with greater degrees of velar elevation and retraction, and of course, with differing tongue, lip, and mandibular positions. This 'old' information is reiterated to point out that it took considerable time and research before the diagnostician could be confident in identifying aberrant tongue and velar postures for the respective vowels produced.

Physiological Events Related to Auditory Phenomena. The physiologic interpretation of cephalometric (still) x-ray films depended, in large measure, upon the specific vowel produced. Similarly, confident interpretations of cine films depends to a great extent, upon the specific utterance recorded. Regardless of the number of frames obtained per second, each frame still records activity occurring within one brief cross-section or segment of time. Many of the physiologic events occur very rapidly and are understood inadequately. For these reasons, the observed physiological activity should be related in time to some aspect of the associated acoustical or auditory event.

Acoustically and physiologically, speech creates a pattern of almost continuous change which is extremely complex and variable. These fundamental characteristics suggest research design which provides for synchronization of sound and film as well as informed phonetic structuring. It follows that careful formulation of the fundamental areas to be studied should precede, not follow, the accumulation of cine records.

Acoustical Specifications of Speech Suggest Formats for Cine Research. The sequential order of research which has emerged in the acoustical specification of speech suggests a format for pursuing the physio-

logical specification of speech. The pattern which has been followed in acoustical phonetic research is recommended for two basic reasons: a) It has been dictated by fundamental principles in phonetics and b) The spectral information provided is accessible and can be utilized to interpret and correlate with the detailed physiologic descriptions developed through cine analysis.

Vowels, as a group, have been studied extensively. However, in addition to the well-defined acoustical parameters of vowels, there are physiological features of vowels which render them easily and reliably identifiable in continuous speech sequences. For these reasons, vowels, selected judiciously, may be considered the anchor sounds or positions for sequential and controlled construction of syllables (*19*). Various consonants, differentiated by physiologic features and combined with vowels, can then be studied as syllabic units. By long term planning, a descriptive repertoire of coordinated speech activity may then be developed. This type of information would be extremely useful in identifying defective coordination, determining which structures, and which muscles require re-education, and in devising effective methods of re-education.

Cine Data With Broad Area of Visualization. Fundamentally, the speech diagnostician is interested in causative factors related to defective articulation. Identification of movement sequences which result in defective sound production may contribute to understanding causative factors as well as suggest remedial procedures. Research, which defines normal speech coordination in its simplest form (syllables) and which is broad in areas of analysis (not restricted to velar movements), will contribute to a physiological specification of normal speech coordination which has broad clinical implications. These implications are not restricted to diagnosis and treatment of disordered speech of the cleft palate patient.

Broad areas of cine visualization and analysis are needed particularly in study of cleft palate speech for two basic reasons: a) Not all speech deviations are attributable to palatopharyngeal deficiency and b) The generation of nasalized speech has two essential features: the influence of the nasal passages upon the spectral output, and the characteristic condition of the vocal tract proper and its changes as a function of time. These facts indicate that cine analysis might be expanded judiciously to include the study of the vocal tract proper as well as the palatopharyngeal region.

Rationale for Standardization of Cine Procedure

At the present time, research and diagnostic use of cineradiography as applied to speech closely approximate each other. For this reason, it is recommended that investigators control and maintain a standardized cine speech procedure when obtaining diagnostic records. Such data over a period of years would then constitute a reservoir of data extremely

valuable for research purposes. Longitudinal analyses of such data would be possible and especially valuable for cleft palate speakers who are heterogeneous in many respects—degree and extent of cleftness, rehabilitative procedures, etc. *With appropriate cine standardization, diagnostic data with research value could be accumulated with minimal expense in roentgen dosage, time and dollars.*

One problem in analysis of cine films is that of selecting from many possibilities the most significant features for study. Essentially, there is too much rather than too little information recorded on the cine film. Restriction in areas of study is a necessity; as a consequence, there may be considerable neglect and wastage of potential information. By standardizing cine methodology for speech study, individual investigators might focus analysis upon differing aspects, and differing phonetic sequences. Thus, an increase in focal areas studied could be realized and a greater variety of coordinated patterns of activity could be defined and comparatively analyzed. Since the surface has scarcely been scratched in cineradiographic study of speech, plans to avoid duplication of research effort and expense might well be considered critically.

Focus: Dynamic Features of Speech

In broad outline, cineradiography should make its major contribution to speech science by providing that aspect of study which has been inaccessible previously. Dynamic features of speech physiology which involve analysis of structural movements within references of space and time appear the logical focus for study. Such study is encouraged by recent work in acoustic phonetics, which has been directed to study of transitions and temporal characteristics of speech.

The transitions are the rapid shifts in the frequency positions of vowel resonances which occur where vowels and consonants join. The rapid movements of speech structures from one position to another are responsible for the shifting resonance response of the vocal tract system. It is believed that the physiologic movement usually produces an equally rapid shift in acoustic output (*2*). Thus, the duration of the acoustic transition between sound is considered an approximation of the time interval associated with physiologic movement from one sound to the next. For this reason, information pertaining to the timing of articulatory behavior might be inferred from study of the acoustical pattern.

Research has shown that the duration of transitions varies according to phonetic classification. For example, transitions are longer in duration when vowels are blended with glides than when the same vowels are blended with plosive and nasal consonants (*12, 17*). The effect of transition duration upon perception also varies in relative importance (*11, 15*). For some consonant-vowel combinations, the duration of the transition has been found to be very important to accurate identification of the sound. For other consonant-vowel combinations, the details of the transi-

tion, including duration, have little effect upon the perceptual identity of the sounds produced. Fricatives, as a rule, do not need support from the transitional sound interval to be identified (*3*).

Research data pertaining to transitions suggest several physiological implications. A difference in speed of speech movements may be indicated by differences in the duration of transitions. Faster movements may occur for some sounds than others. If this assumption is valid and the movement is responsible for the transition, then psychoacoustic investigations suggest that the parameter of speed of movement has a variable effect upon consonant perception which depends partly upon which consonant is produced. Precise timing of speech movements may be critical for identifiable production of some sound groups; for other sound groups, the speed of the movement may be less important to perception.

Lehiste and Peterson (*9*) have stated that the differences in duration of transitions do not appear closely related to inherently different rates of movement for respective speech structures. The same speech structure, such as the tongue tip, may be active in production of several sounds with different temporal characteristics. The difference in duration of transitions were interpreted by Lehiste and Peterson as appearing to be related to differences in manner of production, and to differences in phonetic context. As applied to the discussion of cineradiography, these observations again indicate that phonetic structuring should be considered carefully in developing speech utterances for study.

THEORETICAL CONSIDERATIONS OF PALATOPHARYNGEAL BEHAVIOR AND TRANSITION DURATIONS. The duration of transitions between nasal consonants and oral vowels are particularly interesting in regard to palatopharyngeal function. Theoretically, the time interval of the acoustic transition should be related in some manner to the physiologic shift of the velum from a partially relaxed to an elevated and retracted position. The transition between nasal consonant and vowel (20 to 50 msec) has been described as approximately the same duration as the transition between plosive consonant and vowel (*10*, *13*, *15*). These data suggest that resonance shifts of the vocal tract are about equal in timing even though one consonant may involve a rapid shift in velar position, and the other, no appreciable shift in velar position, per se.[1]

Acoustically defined transitions and temporal aspects of articulatory behavior are fascinating, but constitute hazardous theoretical considerations. Currently, temporal relationships between acoustical and physiological aspects of articulatory behavior are poorly defined. Furthermore, physiological data and acoustical data do not always correlate as expected. Cineradiographic data supplied by Bjork (*1*) and Nylen (*16*) have shown that the time interval required for opening and closing the

[1] Specifically, the duration of transitions for nasal consonant-vowel combinations was found to average 45 msec for /m/; the duration of transitions for voiced plosive-vowel combinations averaged 51 msec for /b/ (*9*).

palatopharyngeal port averaged 120 to 150 msec, respectively. This time specification is considerably longer than that suggested by acoustical study of nasal consonant-vowel transitions.

Theoretical explanations may be postulated for the differential in physiological and acoustical specification of time related to palatopharyngeal activity. The timing of resonance shifts per se may not correlate exactly with complete closure of the palatopharyngeal port. It is also known that the resonance shifts in the spectrum are not the result of changes in the palatopharyngeal region exclusively. The entire vocal tract functions physiologically and acoustically as a system. With this concept in mind, it may be that the timing of the nasal consonant-vowel transitions reflects very rapid movements of structures other than the velum. Velar movements may be slower than movements of other structures which are taking place simultaneously. Recent experimentation in speech synthesis (*5*) tends to support this latter concept.

The present status of the information is one which encourages more questions rather than answers. Generally, investigators seem hesitant to identify exactly which structures, which movements or combinations of movements are responsible for discrete aspects appearing within the spectrum.

To generalize, rapid progress in acoustic phonetics has had a sobering effect in that speech scholars recognize a pressing need to study speech physiology with improved instrumentation. Cineradiography with synchronous sound recording offers a direct approach to improve understanding of changing vocal tract dimensions and the associated change in spectral output within the temporal reference.

Filming Speed. In an organized search for the desired information, an appropriate camera speed is a highly important consideration. The duration of individual speech sounds is known to differ markedly. In slow speech, individual sounds are said to range from 20 msec to 300 msec (*6*). However, the extremely short duration of some sounds, 8 msec and 12 msec for /b/ and /d/ respectively (*3, 7*), and the short duration of transitions indicate a need for camera speeds exceeding 24 frames per second. Moll's cineradiographic study (*14*) has shown that essential physiologic details, such as lingua-alveolar contacts, can be missed at 24 frames per second. Bjork (*1*), who recorded at 50 frames per second, noted that the velum moved as much as 3 mm during the interframe interval. Whereas this degree of movement may not be critical for some studies, it may represent a significant loss of detail in appraising palatopharyngeal function.

When rapid articulatory movements occur, blurring of the image may be expected if there is a relatively long interval of shutter opening. Although the degree of blurring may not be recognized in superficial viewing of the film, it has been found to become very apparent when tracings are made of selected frames for quantitative time-space analysis.

Camera speed, as other details in technology and data reduction, should

be determined by the parameters selected for investigation. After selection of parameters, questions pertaining to appropriate instrumentation arise. If such questions are not asked and answered before accumulation of data, investigators may continue to work with limitations, no longer technologically imposed, but rather self-imposed. As a consequence of scientific and technological complexity, active interprofessional cooperation seems requisite to an intelligent course of physio-acoustic research.

Eastman Dental Dispensary
800 Main Street East
Rochester, New York 14603

References

1. Bjork, L., Velopharyngeal function in connected speech. *Acta Radiol.*, suppl. *202*, 1961.
2. Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J., Some experiments on the perception of synthetic speech sounds. *J. acoust. Soc. Amer.*, *24*, 597–606, 1952.
3. Fant, G., Acoustic analysis and synthesis of speech with applications to Swedish. *Ericsson Technics*, *1*, 1959.
4. Fujimura, O., Bilabial stop and nasal consonants: a motion picture study and its acoustical implications. *J. speech hearing Res.*, *4*, 233–247, 1961.
5. Hecker, M. H. L., Studies of nasal consonants with an articulatory speech synthesizer. *J. acoust. Soc. Amer.*, *34*, 179–188, 1962.
6. Heffner, R-M. S., *General Phonetics*. Madison, Wisconsin: Univ. Wisconsin Press, 1960.
7. Hoffman, H. S., Study of some cues in the perception of the voiced stop consonants. *J. acoust. Soc. Amer.*, *30*, 1035–1041, 1958.
8. House, A. S., On vowel duration in English. *J. acoust. Soc. Amer.*, *33*, 1174–1178, 1961.
9. Lehiste, I., and Peterson, G. E., Transitions, glides, and diphthongs. *J. acous. Soc. Amer.*, *33*, 268–277, 1961.
10. Liberman, A. M., Delattre, P. C., Cooper, F. S., and Gerstman, L. J., The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychol. Monogr.*, *68*, 1–13, 1954.
11. Liberman, A. M., Delattre, P. C., Gerstman, L. J., and Cooper, F. S., Tempo of frequency change as a cue for distinguishing classes of speech sounds. *J. exper. Psychol.*, *52*, 127–137, 1956.
12. Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C., The discrimination of speech sounds within and across phoneme boundaries. *J. exper. Psychol.*, *54*, 358–368, 1957.
13. Liberman, A. M., Ingemann, F., Lisker, L., Delattre, P. C., and Cooper, F. S., Minimal rules for synthesizing speech. *J. acoust. Soc. Amer.*, *31*, 1490–1499, 1959.
14. Moll, K. L., Velopharyngeal function in connected speech. *J. speech hearing Res.*, *3*, 227–241, 1960.
15. Nakata, K., Synthesis and perception of nasal consonants. *J. acoust. Soc. Amer.*, *31*, 661–666, 1959.
16. Nylén, B. O., Cleft palate and speech. *Acta Radiol.*, Suppl *203*, 1961.
17. O'Connor, J. D., Gerstman, L. J., Liberman, A. M., Delattre, P. C., and Cooper, F. S., Acoustic cues for the perception of inital /w.j.r.l/ in English. *Word*, *13*, 25–43, 1957.
18. Powers, G. R., Cinefluorographic investigation of articulatory movements of selected individuals with cleft palates. *J. speech hearing Res.*, *3*, 59–69, 1962.
19. Stetson, R. H., *Bases of Phonology*. Oberlin, Ohio: Oberlin College Press, 1945.