# Synchronous Recording of Speech with Associated Physiological and Pressure-Flow Dynamics: Instrumentation and Procedures

J. DANIEL SUBTELNY, D.D.S., M.S.
ROBERT M. McCORMACK, M.D.
JOANNE D. SUBTELNY, Ph.D.
JOSEPH H. WORTH, M.S.
LESTER M. CRAMER, D.D.S, M.D.
JAMES C. RUNYON, M.S.
ROBERT M. ROSENBLUM, D.D.S., M.S.
*Rochester, New York*

The coordination involved in speech production determines the physical properties of the vocal tract as a modifiable sound-generating and transmitting apparatus. Physiological events are responsible for rapid modifications in the pressure-flow dynamics of the breath stream during speech. The manner in which air pressure and flow are controlled, directed, and emitted determines to a great extent the spectral features of the output and perceptual response. For these reasons, pressure-flow data may be related in a meaningful manner to acoustical features of the output and thereby contribute to understanding the multiple articulatory movements responsible for discrete features of speech.

Within this broad reference, cineradiographic study combined with a study of pressure-flow dynamics of the breath stream and speech output was undertaken to gain a better comprehension of the total speech process. Specifically, pressure-flow dynamics were studied to facilitate the understanding of the palatopharyngeal mechanism which serves to control and direct the breath stream and to modify intraoral pressure during speech. The instrumentation and procedures described here were developed to study normal and abnormal palatopharyngeal valving and the associated physical and perceptual parameters of speech. It is emphasized, however, that the equipment and procedures used in data accumulation and reduction are applicable to a broad variety of research projects.

Studies of single parameters, such as the acoustical characteristics,

---

Drs. Subtelny and Mr. Worth are associated with the Eastman Dental Center in Rochester, New York; Dr. McCormack is associated with the University of Rochester School of Medicine; Dr. Cramer is associated with Temple University Medical School, Philadelphia; Mr. Runyon is associated with the Rochester Institute of Technology; and Dr. Rosenblum is in private practice in Rochester, New York.

This work was supported by PHS Research Grant DE-01837, National Institute of Dental Research.

physiological activity, air flow, and air pressure during speech have and will continue to contribute needed information. However, definition as well as interpretation of independent and interdependent parameters of speech are best facilitated by investigations which provide for the simultaneous observation of these parameters. Such an investigation requires complex instrumentation making it possible to record simultaneously observations regarding: (a) the physiological features of speech activity; (b) the physical characteristics of oral and nasal air pressure, air flow, and sound power; and (c) the speech output.

The purpose of this paper is to describe instrumentation and procedures utilized in obtaining cineradiographic films which are synchronized with recordings of speech, oral and nasal air pressure, oral and nasal air flow, and sound power. The type of data recorded and methods of analysis employed to extract information are illustrated and discussed.

## Instrumentation

In overview, the instrumentation system used to record the desired parameters of speech includes: various transducers for converting sound, air pressure and air flow into electrical signals; high speed cineradiographic equipment; a synchronization system; a stereo tape recorder; a multi-channel data tape recorder and a photographic oscillograph which permits all recorded signals to be written in synchronism on recording paper. The flow of information through the system is presented diagrammatically in Figure 1. As shown, the electrical signals of the sound and air flow recordings receive further treatment by analogue computer techniques to yield sound power and volumetric data for oral and nasal flow.

Cineradiographic Equipment. The equipment for synchronized cephalometric cineradiography consists of: x-ray generating equipment, an image intensifier assembly, a high speed 16 mm motion picture camera, synchronizing circuits, a TV vidicon monitor system, a cephalostat and associated mounting supports. These components are housed in a control room and in an adjoining sound treated room equipped for satisfactory intercommunication.

A modified Wehmer Cephalostat is used to orient and stabilize the subjects. Four adjustable rubber cups and the nasion rest are used to stabilize the head after proper positioning has been established by using the conventional ear rods. The ear post assemblies are removable to lessen interference with auditory feedback and thus, the articulatory activity of the subject.

The x-ray source is a General Electric DXR-1650 constant potential six-phase generator. The timing circuits of this x-ray unit are synchronized with the output of a pulse generator within the motion picture camera. The motion picture camera is a Cinerama Monitor 500

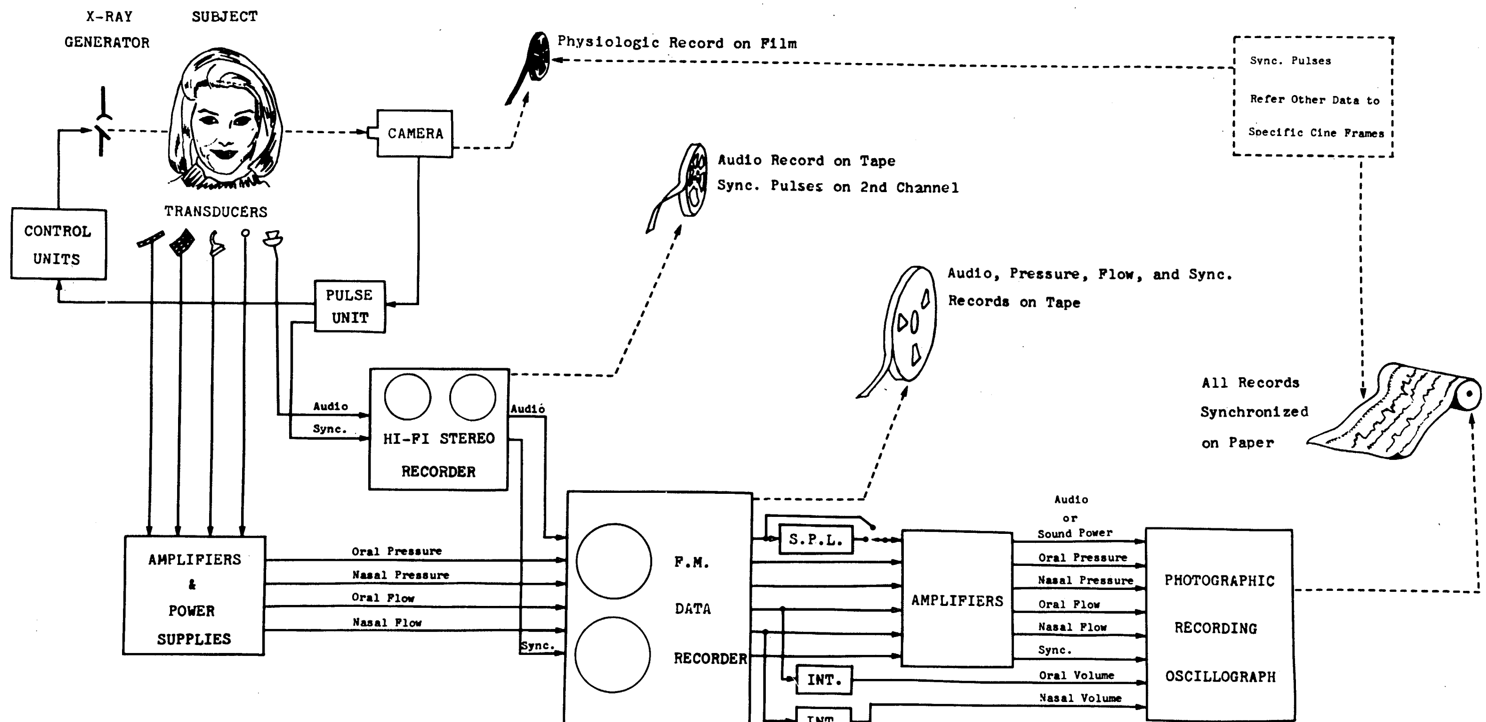


FIGURE 1. Functional diagram of the instrumentation system. Simultaneously, with the cineradiographic exposure, speech-connected parameters are converted into electrial signals and recorded on magnetic tape. Cine-synchronizing pulses are also recorded, as illustrated. Later, these data are played back, treated by analogue methods, and written out as a synchronized graphical record.

with an intermittent, pin-registration movement. Although capable of operation up to 500 frames per second, the camera is used at a rate of 240 frames per second for present purposes.

Noise generated inside the sound-treated room by the motion picture camera is reduced approximately 18 dB by blimping.[1] Since the camera must remain fixed to the image intensifier and television chassis, some audible noise is conducted to these units. For this reason, both units, excepting the face of the intensifier, are covered with a blanket made of glass wool bonded to vinyl.

The image intensifier system utilizes a 9″ tube with a 6,000 gain. The intensifier assembly is mounted on a sliding carriage for precise movement along the optical axis. The distance from the anode of the x-ray tube to the mid-sagittal plane of the subject is fixed at 27 inches. The additional distance to the image intensifier face is maintained at a minimum of 7 inches. The minified image on the output phosphor of the intensifier tube is recorded by the motion picture camera and is monitored simultaneously by the television system.

Proper exposure, ranging from 90 to 98 killivolts, is determined by two measures of facial width. These measures are made at a point immediately posterior to the mandibular gonial angles and at the most prominent points of the zygomatic arches. A fixed setting of 1.30 milliamperes is used with Kodak Plus X Film.

Because pulses generated by the cine camera are used to drive the x-ray tube, the x-ray beam at 240 frames per second is on for only 24% of the time. Thus, this system reduces radiation dosage considerably. As measured with a condenser R. Meter, a radiation skin dose of 2.7 roentgens per minute was recorded. During an average recording sequence of 20 seconds, subjects receive about .9 roentgens.

Synchronization System. Synchronization between the cine camera and the x-ray circuits is accomplished by the pulse control unit. The shutter opening is 130°. As shown in Figure 2, a correlation pulse circuit within the camera generates one pulse per frame. This correlation pulse is generated 25° before shutter opening and is conducted to the pulse control unit where it is delayed .44 msec and increased in duration to 1 msec. The pulse is then conducted through the x-ray control unit to the generator timing circuits which switch the anode voltage to the x-ray generator tube. Thus, the x-ray pulse is initiated .15 msec after the camera shutter starts to open. The purpose of the variable pulse delay is to allow optimum timing of the x-ray pulse in relation to the cine camera shutter opening.

---

[1] The blimp was constructed according to directions obtained from Sonex Cinema Engineering Company, Glendale, California, for the use of their "Hydro-Seal" lining process (Patent applied for).

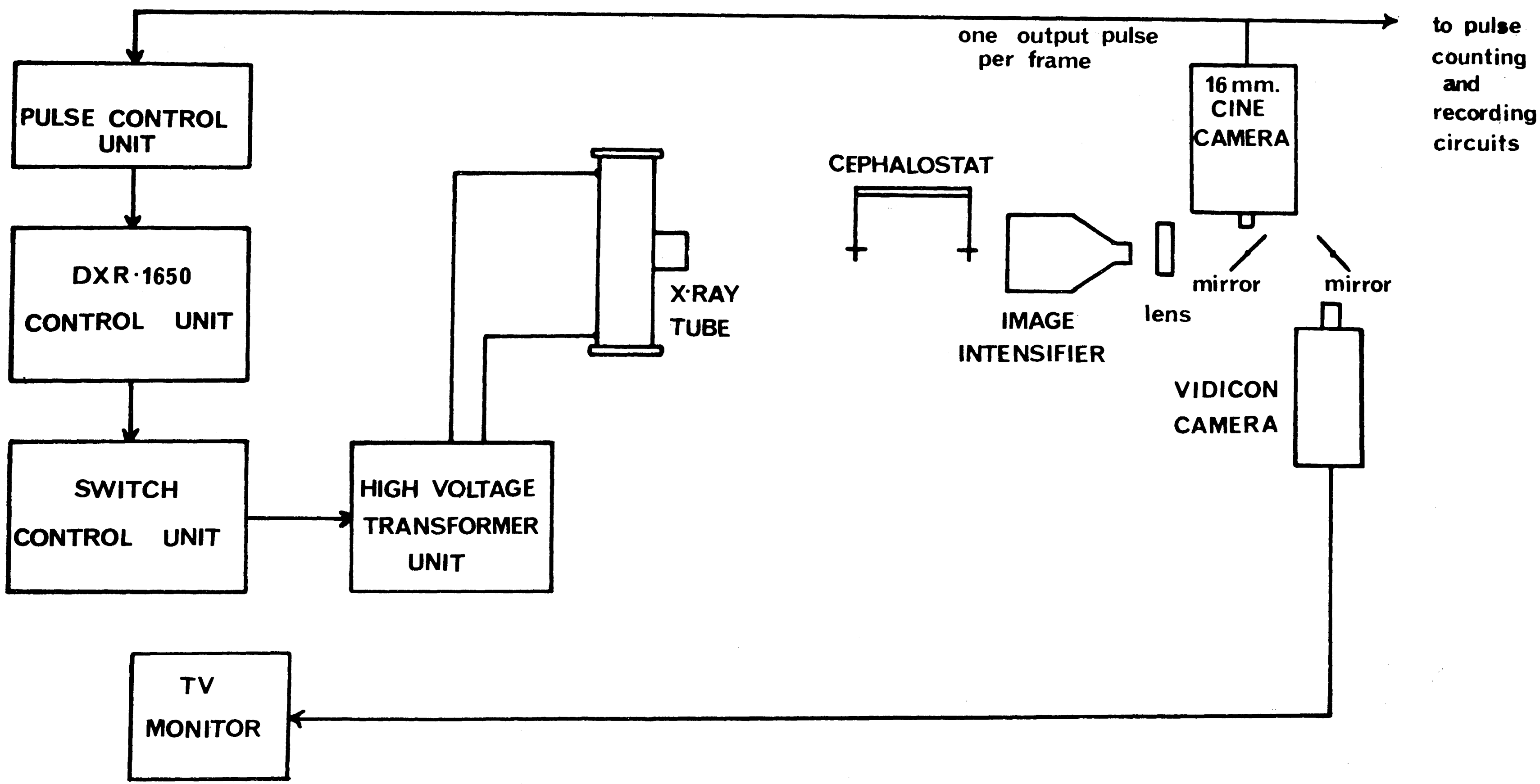


FIGURE 2. Functional diagram of the equipment for high-speed synchronized cephalometric cineradiography. The path of the synchronizing pulses generated by the cine camera is shown. Pulses are used to drive the pulsed x-ray tube, and are also recorded to synchronize the cineradiographic record with all other data.

At 240 frames per second, the interframe interval is 4.17 msec; the shutter opening is 1.5 msec and the x-ray pulse duration is 1 msec. Thus, the cine film records one millisecond "slices" of articulatory activity recorded at 4.17 msec intervals.

The system providing for synchronization of cineradiographic and pressure-flow data is diagrammatically included in Figure 1. As shown, the pulses generated by the cine camera are used for two purposes. The first synchronizes the x-ray tube with the shutter opening as described; the second synchronizes the cine film with other data. For the latter purpose, pulses generated within the cine camera are fed to a Weston 2007 divider for modification to produce a series of decade-scaled pulses. The modified signal is then recorded on one channel of the stereo tape recorder so that the speech recorded on the other channel can be synchronized with other parameters in future spectrographic analysis if desired.

For present purposes, the synchronizing pulses recorded on the stereo tape recorder are taken from a monitoring output and recorded on one channel of the data tape recorder. When written out through one channel of the oscillograph, the pulse train has every 10th, 100th, and 1,000th pulse clearly distinguishable. This form of pulse write-out facilitates identification of a specific cine frame which occurred simultaneously with some feature of audio, pressure, or flow data. A Kodak Analyst 16 mm movie projector with selective frame advance and counter is used to locate specific frames of interest.

Compensations for magnification and minification factors are made on the basis of the projected image size of the intraoral pressure transducer positioned in the mid-sagittal plane. The Benson-Lehner Oscar Model F with decimal converter is used for film analysis. This instrument allows control of the write-out scale so that all measurements may be recorded directly in the desired units.

TRANSDUCERS FOR RECORDING SOUND, AIR PRESSURE AND AIR FLOW. The sound produced by the subject is detected by an AKG D19-E microphone maintained six inches from the lips at an angle depressed 45° from the horizontal. Audio signals are recorded on one channel of an Ampex F-44 stereo tape recorder for use in speech evaluation and sound identification. The audio signals are also taken from the monitoring output of the Ampex for recording on one channel of a Honeywell 8100 8-track frequency-modulated data tape recorder.

*Pressure Recording.* The transducers, their calibration and use for recording of intraoral and intranasal pressure, are described elsewhere by Subtelny, Worth, and Sakuda (*3*). In brief, the recording of intraoral pressure is accomplished by a small silicon strain gauge pressure transducer pasted to the roof of the mouth at a midline point slightly posterior to the rugae. For intranasal pressure recording, a short catheter cemented to a similar transducer is inserted into one nostril. The frequency response of both transducers (±1.5 dB from 0 to 400 Hz) is more than adequate to record pressure modifications caused by articulatory movements and, furthermore, permits recording of very rapid pressure fluctuations caused by vocal fold vibration (Figure 3).

Intraoral and intranasal pressure signals are appropriately amplified and recorded on a second and third channel of the data tape recorder. The pressure signals from the data recorder are then fed through the Honeywell T6GA current amplifier to the photographic oscillograph (Honeywell 1108 Visicorder). In write-out, oral and nasal pressure recordings are scaled so that pressures of 1 cm $H_2O$ are recorded by deflections of 1 and 3 chart divisions respectively. Time is scaled on the horizontal with 1 mm equal to 5 msec.

*Flow Recording.* The apparatus for detecting and recording oral and nasal flow as electrical signals was designed and constructed to eliminate the use of a face mask to trap air exhaled from the nose and mouth. Fundamentally, the flowmeter permits simultaneous recordings of flow, audio and pressure in a manner which does not interfere with normal articulatory movements or the quality of sound produced. The instrument, its calibration and performance, has been described in detail by Worth and Runyon (*5*).

To generalize, the flowmeter consists of a series of short warm-wire sensors arranged in oral and nasal sections which are separated by a thin metal plate. In use, the metal plate divider is positioned to contact the upper lip at the vermillion border. The oral section, consisting of 24

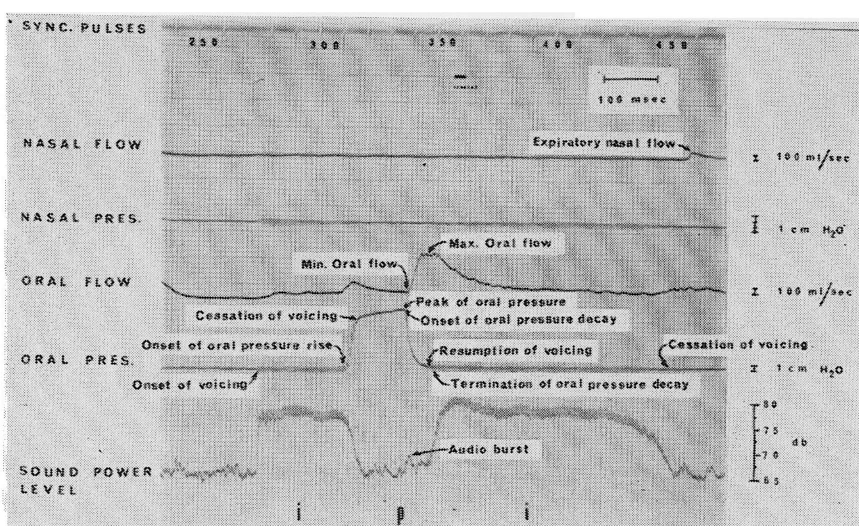


FIGURE 3. Physical events utilized in data analysis are identified in the above synchronous recording. The heavy and dashed lines appearing on the horizontal axis indicate respective durations of 21 msec and 42 msec which would be corresponding intervals defined by camera speeds of 48 and 24 frames per second.

warm wires, samples the velocity of flow at uniformly spaced intervals in a curved area around the mouth. Each sensor is served by a separate feedback amplifier and linearizer. The linearized velocity signals are added by an operational amplifier used as a summer. The gain of the summer is adjusted to obtain a desired sensitivity.

The nasal section of the flowmeter consists of four warm wire elements. Unfortunately, four sampling points for recording nasal flow were found to be inadequate. As a result, calibration of the nasal section is necessary for each subject. Calibration is accomplished by having each subject produce tidal volumes, first orally and then nasally. These volumes are recorded and integrated. The gain of the nasal summing amplifier is then adjusted to give the nasal section of the flowmeter the same sensitivity as the oral section.

Primarily because of the inadequate number of nasal flow sampling points, angular positioning of the meter relative to the nostrils of the speaker is critical to the reliability of *quantitative* analysis. These factors, which have been experimentally studied, indicate that nasal flow is recorded less reliably than oral flow.

Oral and nasal flow signals are recorded on two additional channels of the data tape recorder and are written out on separate channels of the oscillograph at a scale of 100 ml/sec for each 2 mm chart division.

Analogue Computer Treatment of Flow Data-Volume. In order to obtain the volume of oral and nasal air displaced as a function of time, the electrical signals representing the volume velocity of oral and nasal air flow are integrated with respect to time. A block diagram of the circuits used for this integration is presented in Figure 4.

The integrator consists of an operational amplifier with capacitive feedback. A bias current is injected at the integrator input in order to establish constant output with zero flow input. The resistor at the in-

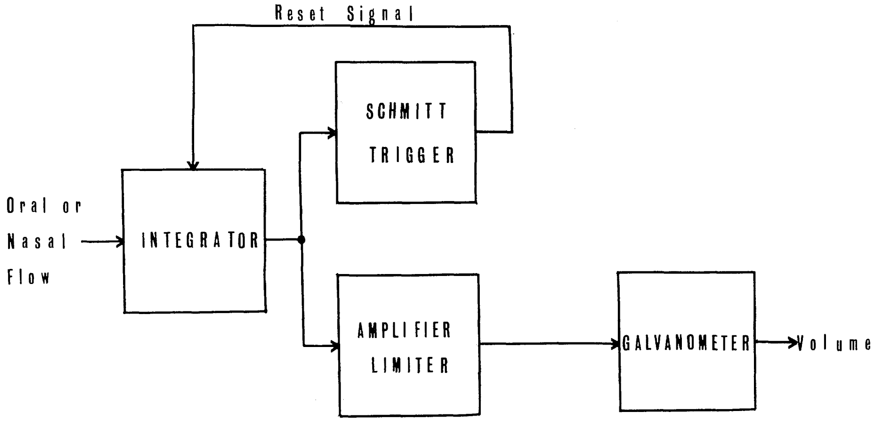


FIGURE 4. Block diagram of the circuits used for integration of flow signals.

tegrator input is adjusted so that a steady flow of 100 ml/sec produces a sensitivity of 10 ml/div on the graphical record. To facilitate analysis of oral flow volume, the trace is made to return automatically to zero or to reset each time output voltage reaches a predetermined voltage less than the saturation voltage of the amplifier. The reset relay is controlled by a Schmitt trigger which senses this voltage.

AUDIO DATA-SOUND POWER. Since both pressure and volume velocity of flow are influenced by the intensity of voice or the degree of vocal effort expended, some measure of sound power is needed as reference for study and interpretation of the related physical features of pressure and flow. To gain recordings of sound power level as a function of time, the audio signals (limited in bandwidth to 5,000 Hz) from the data tape are used.

An adequate system for the sound power conversion was assembled as described by Peterson and McKinney (*1*). The system includes a square law rectifier, a filter, and a logarithmic compressor. The photographic oscillograph, fed by a current amplifier, is used to display the sound power signal with a write-out scale of approximately 2 mm per dB. For accurate measurement of the graphical sound power trace, templates were constructed for each recording level used to secure data.

In summary, the instrumentation described has made it possible to obtain high speed cineradiographic films synchronized with other data recorded simultaneously on the data tape recorder. A completely synchronized graphical record is obtained when the tape is played back into the oscillograph. Events occurring at the same time are written at the same vertical position on the record. Since analogue computations are performed in real time, no distortion in time is introduced by the generation of sound power or integrated flow signals. These signals, therefore, are also produced on the record in complete synchronism.

The recording of a normal utterance of *Sing a song of sixpence* in Figure 5 illustrates the type of data provided. The accuracy of utterances is ascertained by making phonetic transcriptions of the audio played back from the Ampex stereo recorder. Phonetic events are properly identified and designated in time on the graphical record by careful monitoring of the speech output and by various features displayed by the physical and physiological recording system.

**Procedure**

Phonetic structuring of utterances for recording during simultaneous cineradiographic exposure is dictated by three basic criteria: (a) brevity, to curtail duration of exposure; (b) simplicity, to assure that the level of articulatory competence required for satisfactory utterance can be realistically expected in a broad variety of subjects; and (c) pertinence, to provide appropriate data for controlled multidimensional study of a maximal number of discrete phonemes articulated in simple

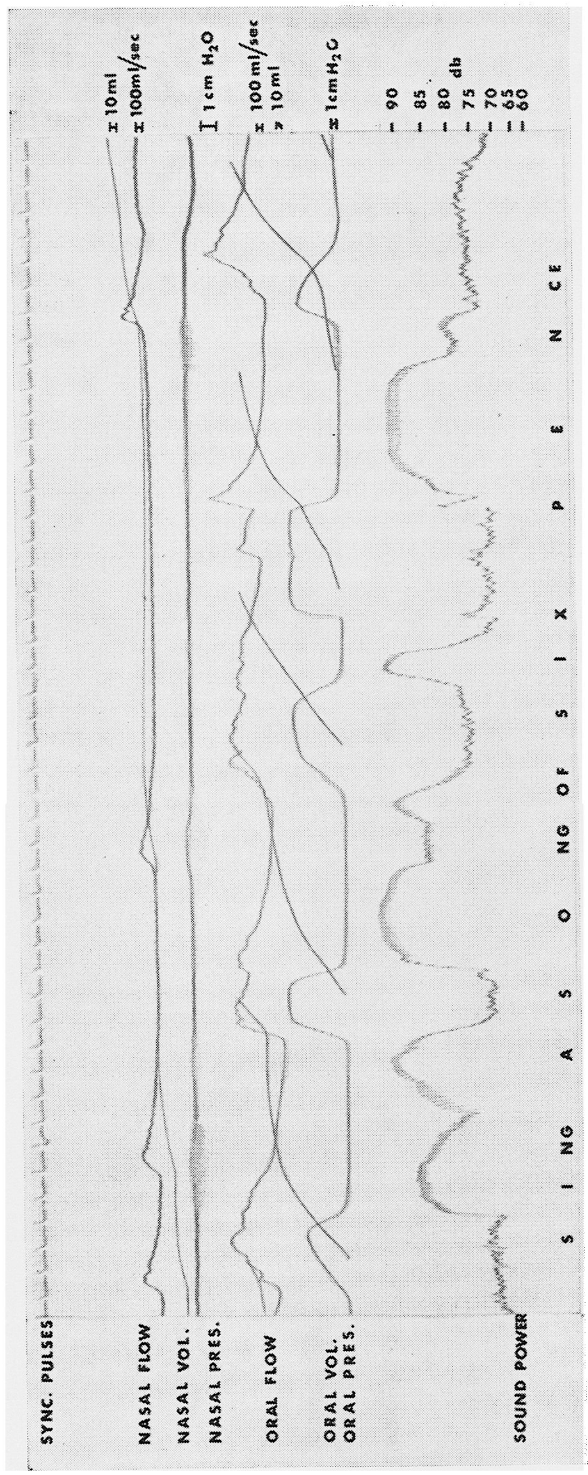


FIGURE 5. Graphical record of the sentence *Sing a song of sixpence*, produced by an adult female speaker.

and controlled phonetic environments and during continuous speech production. For the latter purpose, sentences are structured to provide variety in phonetic environment while still retaining focus on the specific phonemes previously recorded in VCV sequences.

Each subject is given practice in reading the desired speech material before recording. No attempt is made to control pitch or loudness level other than the instruction to use a conversational level of speech. Subjects are also given brief descriptions of the equipment and explanations of the experimental procedure.

After indoctrination and speech instruction, the subject is seated and properly positioned for exposure. Denture adhesive (Orafix) is applied to the back of the oral pressure transducer which is then held in place by digital pressure for about a minute. The tongue tip is marked with a radiopaque mixture consisting of Orabase paste and tantalum powder. The flowmeter is positioned, the nasal catheter inserted, ear posts removed, and the image intensifier positioned as close as possible to the subject in order to reduce geometric distortion (Figure 6).

Before exposure, recordings of test sequences are made and written out in compressed form to adjust the microphone recording level, to balance the circuits and to check recording outputs of all transducers.

**Data Analysis**

Physiological activity during any given utterance appears continuous rather than intermittent in character. Certain events identified at one point in the cine film are not always associated with noticeable modifications in pressure, flow, or sound at that particular point in time. To illustrate, the soft palate elevates and retracts to contact pharyngeal tissue well in advance of sound generation. This observation and others emphasize the fact that physiological events, highly important to the total process of speech production, can occur as anticipatory as well as sequential and synchronous activity. Because of this fact, analysis procedures were developed: (a) to describe physiological events; (b) to specify the timing of events; and (c) to gain measures of pressure, flow and sound power which assist in the interpretation of observed events. To fulfill these objectives, the selection of specific points in time for detailed study is determined by projected film viewing to identify by frame number a particular articulatory event, and by study of the graphical record.

Since physical parameters of speech sounds have a uniqueness in character comparable to their respective physiological and acoustical features, details in measurement procedure are necessarily modified as a function of the specific sound produced and the phonetic environment of the sound. For present purposes, principles and general procedures of analysis are discussed with measurement details described for only two consonants, /p/ and /m/, appearing in constant vowel environments.

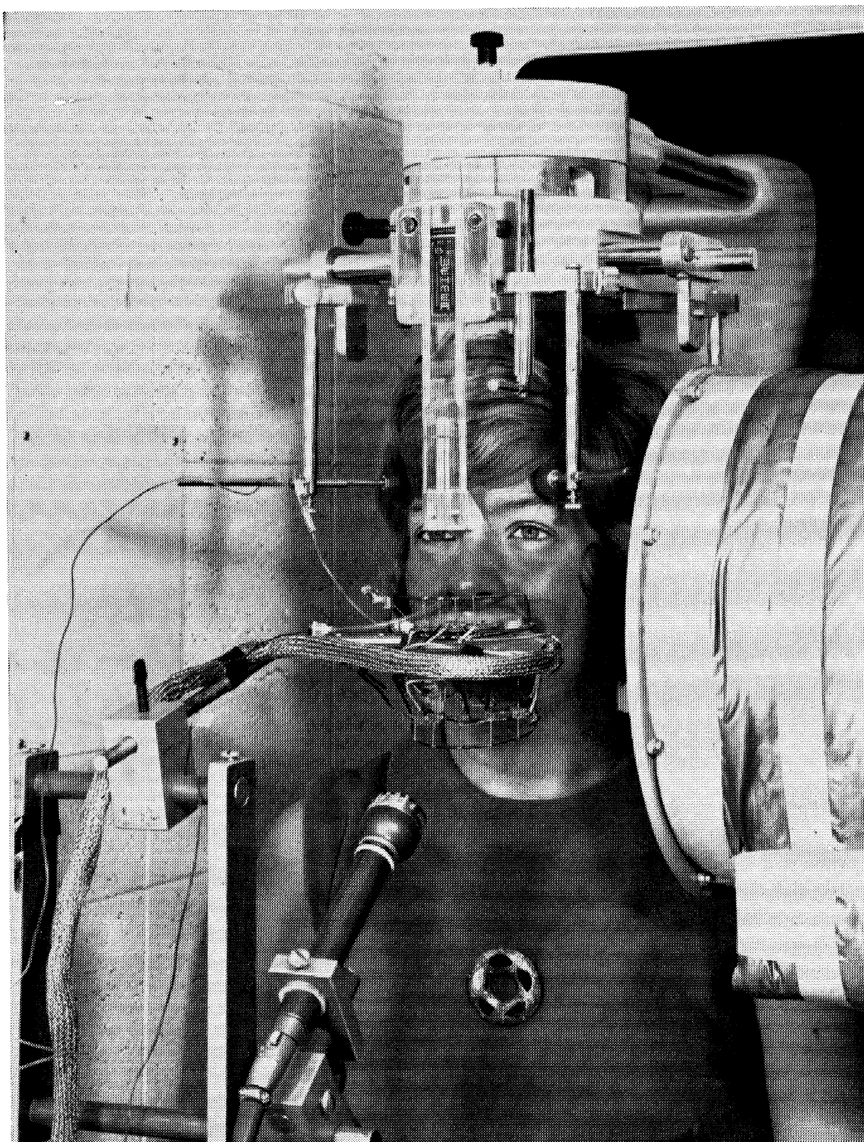

FIGURE 6. Photograph of a subject positioned for recording. The x-ray source is through the port on the left. The input of the image intensifier is visible to the right. Details which are visible include: the cephalostat, location and angulation of the microphone, the flow transducers, the nasal pressure catheter and the flexible lead to the oral pressure transducer.

Data derived by procedures described have been presented for bilabial stop and nasal consonants by Subtelny, Kho, and Subtelny (*2*).

Sound Power Analysis. Noise levels at the recording site in the sound-treated room were evaluated with a Rudmose RA-100 sound analyzer under three conditions: with all equipment off, with equipment ready for recording, and with all equipment including the blimped camera in operation. Noise levels for these three conditions averaged 42 dB (re .0002 dn/cm$^2$), 61 dB, and 65 dB, respectively. Thus, for synchronized recordings, a background noise level of 65 dB is used as reference for SPL analysis. Elevations in SPL above the reference level are measured and interpreted as relative, rather than absolute, measures associated with speech events.

Analysis of ambient noise finalized in a level higher than expected or desired. Despite this fact, the capability of objectifying relative differences in vocal efforts considerably improves the quality of the derived data. To standardize measurement procedure, the SPL trace is bisected providing a reference point from which readings on the appropriate superimposed template are made.

Pressure Analysis. Intraoral pressure recordings yield concise and clearly differentiated shifts in trace as a function of time. In addition, subtle pressure fluctuations caused by the vibratory activity of the vocal folds are recorded. These fluctuations appear as vertical striations conforming to the fundamental or low frequency components of the glottal tone. As would be expected, the signals of glottal activity correspond in time with abrupt shifts in sound power recordings (Figure 3).

Because the pressure trace provides cues of glottal as well as oral valving, the trace is used for measures of oral pressure and as a primary temporal reference. Points are selected to define significant intervals or instants in time for correlation with laryngeal, palatopharyngeal, and oral valving. Intervals defined by oral pressure elevation are also used effectively to differentiate intervals for air flow measurements. Pressure recording for this latter purpose is justified because a complex inter-relationship between pressure and flow is recognized. At present, pressure fluctuations rather than flow patterns can be more confidently and knowledgeably related to physiological events.

Precise measures of amplitude (cm $H_2O$) and duration of intraoral pressure modification are made directly from the trace according to the scale of write out. The insertion of the catheter for a nasal pressure recording does not impede nasal respiratory flow. For this reason, in normal speakers no elevation in nasal air pressure per se would be expected, nor is it detected.

Flow Analysis. Each of the warm wires was calibrated and linearized separately ±5% by comparing its response with that of a Flowtronic 55-Al warm wire sensor, simultaneously exposed to a variable flow source. To calibrate total flowmeter performance, several procedures utilizing mechanically controlled variable flow sources were employed. Reasonable facsimiles of human flow velocity and pattern, in time and space, could not be generated mechanically. A human source of flow generation, therefore, appeared desirable for calibration purposes.

Measurements of vital capacity were made with the flowmeter and with a wet spirometer. By comparing these measurements for each individual, it appears that oral airflows during speech can be measured with a standard deviation of 15% over the normal range of airflow (0 to 1200 ml/sec) with a frequency response down 1.5 dB at 280 cps. On this basis, the flowmeter is considered adequate to indicate volume velocity of flow from the mouth and, with somewhat less reliability from the nose.

The warm wire sensors are nondirectional in their response to air movement. This feature has created no major problem because speech is produced predominantly during the expiratory phase and many physical events recorded by the system accurately define the onset of speech related activities, thereby defining the gross direction of air flow. Reverse flow, however, does sometimes create small artifacts in flow patterns when rapid lip movements occur.

At high velocities of flow, some artifact in recording is also introduced by the entrainment of air within the flow field. That is, individual wires within the flow instrument do respond to a turbulent pattern of flow at very high velocities. This has the effect of causing a slight increase in sensitivity with velocity.

The warm wire response to turbulence in an unimpaired flow field requires consideration in interpretation of measures when high velocities are recorded. The slight distortion in the graphical record of flow vs. time, however, does not detract from the essential information provided by the instrument which reliably records relative changes in flow resulting from articulatory movements. Nonlinearities due to turbulence and other features which can result in some distortion in the graphical record have been discussed and explained by Worth and Runyon (*5*).

Measures of volume velocity are made at generalized levels of consistent flow existing for comparatively long durations of time, as in vowel phonation, and at discrete points in time specified by other events recorded by the system. For this purpose, intraoral pressure and sound recordings are utilized effectively. At designated points on the graphical record, the velocity of flow in millileters per second, is read directly from the flow trace according to the scale of write-out.

No assertion is made that there are no shortcomings in the warm wire flow instrumentation. Indeed, certain shortcomings as mentioned above indicate that further improvement of flow instrumentation would be desirable. However, on the basis of cumulative data, the present flowmetering device seems to fulfill adequately the purposes for *multiple* and *simultaneous* recording of flow, pressure and audio. As described, the flowmeter is recording a relatively inaccessible parameter which is significantly related to both physiological and acoustical aspects of speech.

VOLUME ANALYSIS. Information particularly germane to studies of palatopharyngeal function is provided by volumetric analyses. From the electronic integration of flow over time, the volume is determined for discrete intervals of palatopharyngeal opening or oral port closure as defined by film analysis. Likewise, the volume of oral flow during a voiceless interval of consonant production can be derived for speakers with normal or abnormal palatopharyngeal valving.

To derive oral volume measures, the volume traces are vertically intersected to differentiate the interval of interest. The measure of volume displacement occurring within the differentiated interval is then read according to scale of the integrated flow trace.

## Descriptions and Illustrations of Recorded Data

THE /ipi/ UTTERANCE. For purposes of illustration, synchronous recordings are reproduced in Figure 7. The sixth channel at the top records synchronizing pulses at 4.1 msec intervals with every tenth pulse augmented. The ordering in time of physiological events is illustrated by reproduced tracings of the corresponding frames of the film.

Velar movement from rest (frame 265) to initial touch contact (326) is illustrated in the first two tracings. By studying the ten frame interval, beginning with 387 and ending with 396, the ballistic nature of events during the imploding phase of /p/ is displayed. These events include: the initiation of abrupt rise of oral pressure, reduction of oral flow, cessation of voicing, and diminution of sound power level. These gross shifts in physical parameters occurring within an interval of 41 msec, are related to closure of the oral port and opening of the laryngeal valve with associated cessation of vocal fold vibration.

The third tracing, 387, corresponds in time with initiation of rise in oral pressure. Vocal fold vibration is still in process as registered in the oral and nasal pressure and in sound power recordings. These events are associated with closing movements of the lips and firm velopharyngeal closure.

Continuation of the lip closure, with touch contact established for articulation of /p/, is shown in frame 388. Amplitude of intraoral pressure continues an abrupt rise; sound power level, still registering fold activity, is just beginning to decrease slightly. At this interval, a slight spike in oral flow is recorded. This spike in oral flow recorded at the time of the initiation of rise in oral pressure is attributed to a brief interval of accelerated or disturbed air movement caused by the closing movement of the lips.[2]

At frame 393, vocal fold vibration ceases while bilabial and velo-

[2] With the movement of lip closure, some back flow (or turbulent eddies) of air may occur and be recorded as "flow". This suggested explanation is supported by other observations of increased oral flow coinciding in time with lip movement during bilabial stop production.

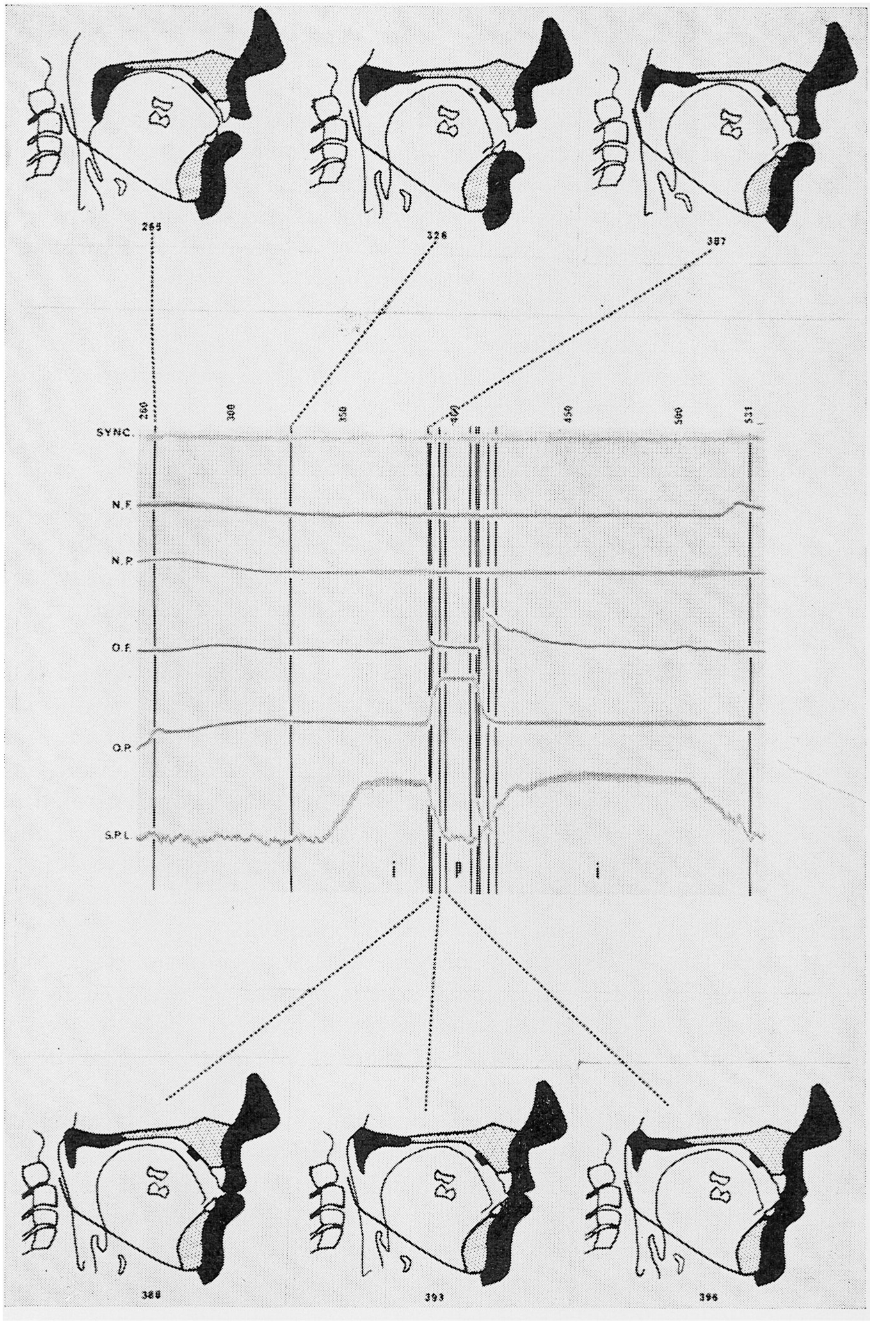


FIGURE 7. Tracings of frames selected from a cineradiographic film which was synchronized with recordings of: sound power, intraoral pressure, oral flow, intranasal pressure, nasal flow and the synchronization pulse. The first two tracings, frames 265 and 326, are reproduced to illustrate in time the anticipatory nature of palatopharyngeal closure. The 10-frame interval (#387 to 396) covers a 41 msec duration phonetically identified as the implosive phase of bilabial plosive production. Physical events associated with the physiological adjustments, defined by respective tracings, are recorded and are identified in proper time relationship by the solid lines intersecting the graphical record.

pharyngeal valves are firmly closed. At this point, amplitude of oral pressure approaches a peak. No nasal flow is recorded and oral flow is at a minimum, not zero, as would be expected during an interval of tight lip closure and high intraoral pressure. The low velocity of oral flow recorded approximates the rate recorded during the preceding vowel produced with associated lip opening. Technically, this finding is explained by the response of the warm wire flowmeter to the exponential decay of turbulence. The rate of indicated decay of oral flow tends to support this explanation.

In frame 396, peak pressure is achieved with oral flow at a minimum and sound power approaching silence. Firm compression of the lips and a high velar position are associated.

In Figure 8, the explosive phase of bilabial plosive production is displayed. Frames 406 and 409 precede the rapid shifts in pressure and flow constituting explosion. The physiological status for the silent interval immediately preceding the spectral burst, characteristic of voiceless stops, is shown in frame 406. The lips remain firmly closed with the amplitude of oral pressure retained at a level of 9 cm $H_2O$. Oral flow is minimal and nasal flow zero, as would be anticipated by anterior closure of the oral passage and posterior closure of the nasal passage.

Peak amplitude of oral pressure terminates in frame 409, immediately preceding explosion. Oral flow remains minimal, no voicing is indicated. The palatopharyngeal valve remains closed while the oral valve has commenced opening.

In the next frame (410), 4.1 msec later, the tongue has arched slightly higher, the lips have opened wider and the palatopharyngeal valve remains closed. Sound power at this point registers a spike, 13.5 dB above the preceding level. The sound burst is of short duration (10 to 20 msec). The spike coincides in time with rapid pressure decay and abrupt rise in oral flow velocity.

The maximum velocity of oral flow (1400 ml/sec) coincides in time (frame 413) with reduced oral pressure (2.5 cm $H_2O$), which is still in the process of decay. Although the burst recorded in sound power level is over, vocal folds at this point have not resumed vibration. Physiologically, little change is associated with these major shifts in physical parameters. The palatopharyngeal port remains closed; the lips open.

In frame 417, vocal fold activity is resumed, pressure has decayed to almost base level, and oral flow has decreased. Gradual decay of sound power is observed to commence at frame 500; however, vocal fold activity continues for a considerable time until zero is attained on the sound power trace. At frame 531, sound power reaches zero and expiratory nasal flow is recorded as the soft palate drops away from the pharyngeal wall.

THE /imi/ UTTERANCE. The temporal character of palatopharyngeal

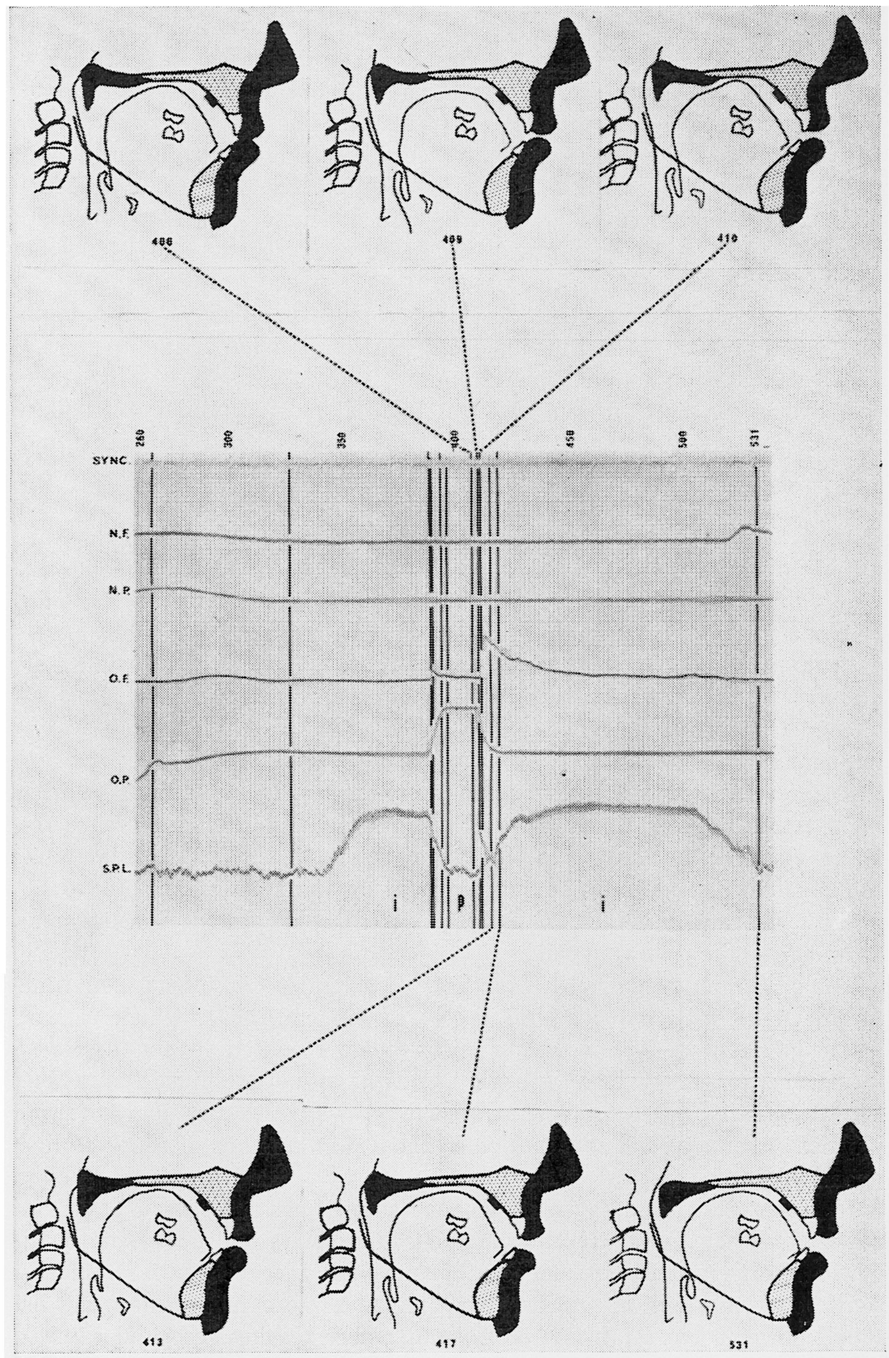


FIGURE 8. The terminal phase of implosion (frames 406 and 409) and the explosive phase of bilabial plosive production is displayed by reproduced tracings of the cine film synchronized with the other recordings of physical parameters.

and oral valving and its effect upon audio and pressure-flow dynamics are displayed effectively in the synchronized recording of the /imi/ utterance (Figure 9). At the juncture between the /i/ and /m/, touch velopharyngeal contact remains (frame 870). At frame 871, velopharyngeal contact is broken, and the shift in audio to nasalization is recorded in sound power, intranasal and intraoral pressure traces. In the next frame (872), the lips initiate closure for the /m/ with a lower and fronted velar position associated. Three frames later (875), lip closure is firm with a co-existing palatopharyngeal aperture approximating 5 mm. Under these conditions, with resistance to oral flow increased, nasal flow velocity begins to rise.

The sequence of physiological and physical events associated with audio shifting from nasal to oral at the second juncture between /m/ and /i/ shows essentially the same relevant facts (Figure 10). Within an interval of 25 msec, pressure, flow and audio parameters are changed significantly as rapid modifications in oral and palatopharyngeal valving occur.

Although temporal features of valving have been found to vary markedly from one speaker to the next, the features of this one subject are described to emphasize the utility of camera speeds exceeding 24 or 48 frames per second. At 24 frames per second, a corresponding 42 msec interframe interval exists which, according to data described, may be too long to record some features. With a camera speed of 48 frames per second, the frame to frame interval is reduced to 21 msec. In the upper portion of Figure 3, the relative durations of these intervals are graphed according to the time scale utilized in present data recording. As shown, some detailed information may be lost or obscured at slow camera speeds particularly when the frame to frame interval coincides in time with implosion or explosion of stop consonants, or with transitions between sounds.

**Discussion**

Sequential, anticipatory, simultaneous, and overlapping are all appropriate adjectives which have been employed to describe articulatory events. Despite description, comparatively little research has been directed toward the temporal specification of events. Synchronous recordings of audio, pressure-flow dynamics and the physiological events filmed at 240 frames per second permit combined study of physiological activity within temporal references and the associated physical parameters which are pertinent to sound generation. Such study is needed to improve understanding of normal speech production and to define the degree of variability expected within any normal reference.

In working toward a normal reference, unique characteristics of respective phonemes and subtle variations in phonemic features attributed to differences in phonetic context and linguistic structure must be

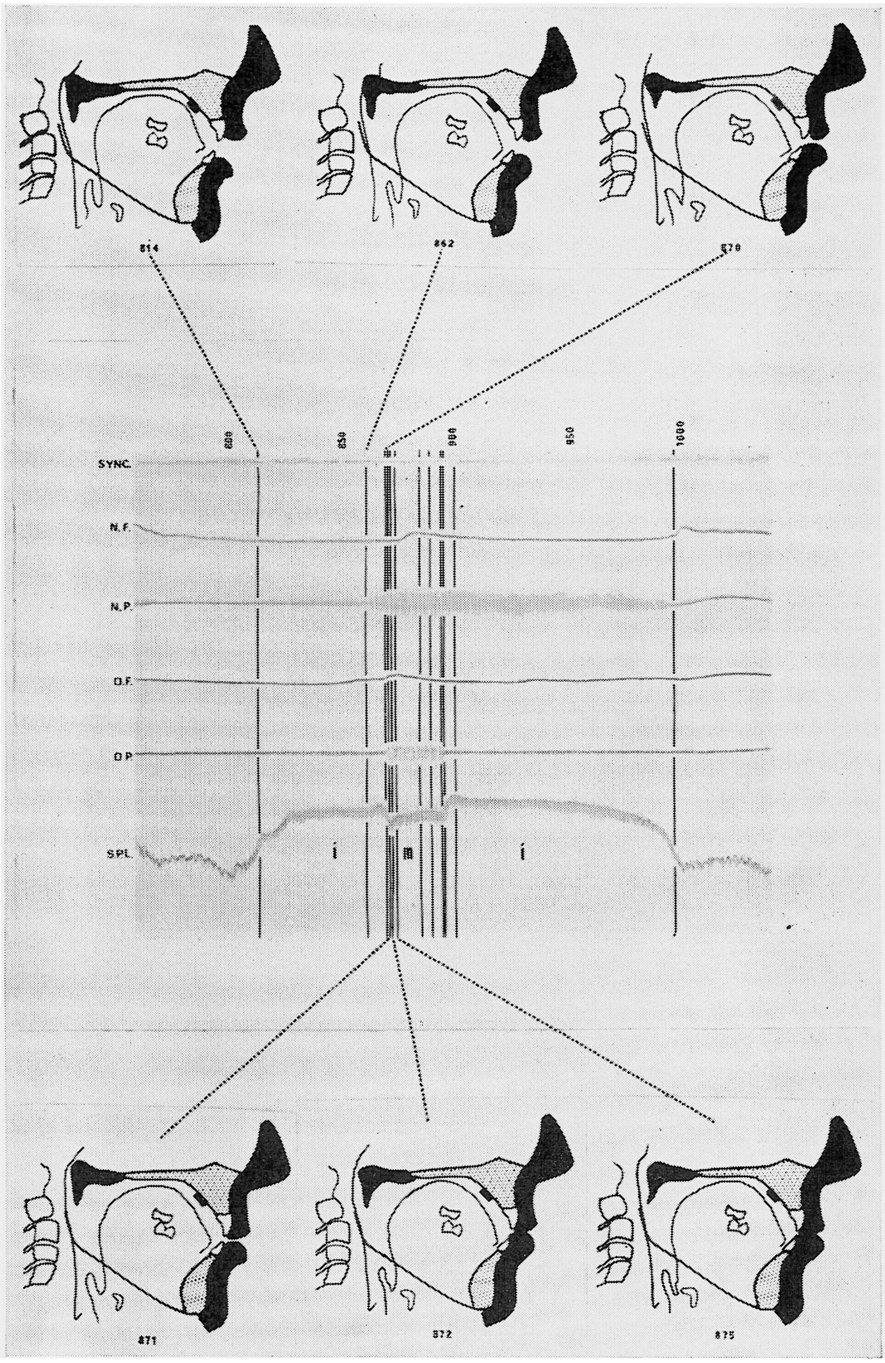


FIGURE 9. The tracings above are reproduced to display the temporal character of physiological events associated with the recorded transition between /i/ and /m/ in /imi/ articulation.

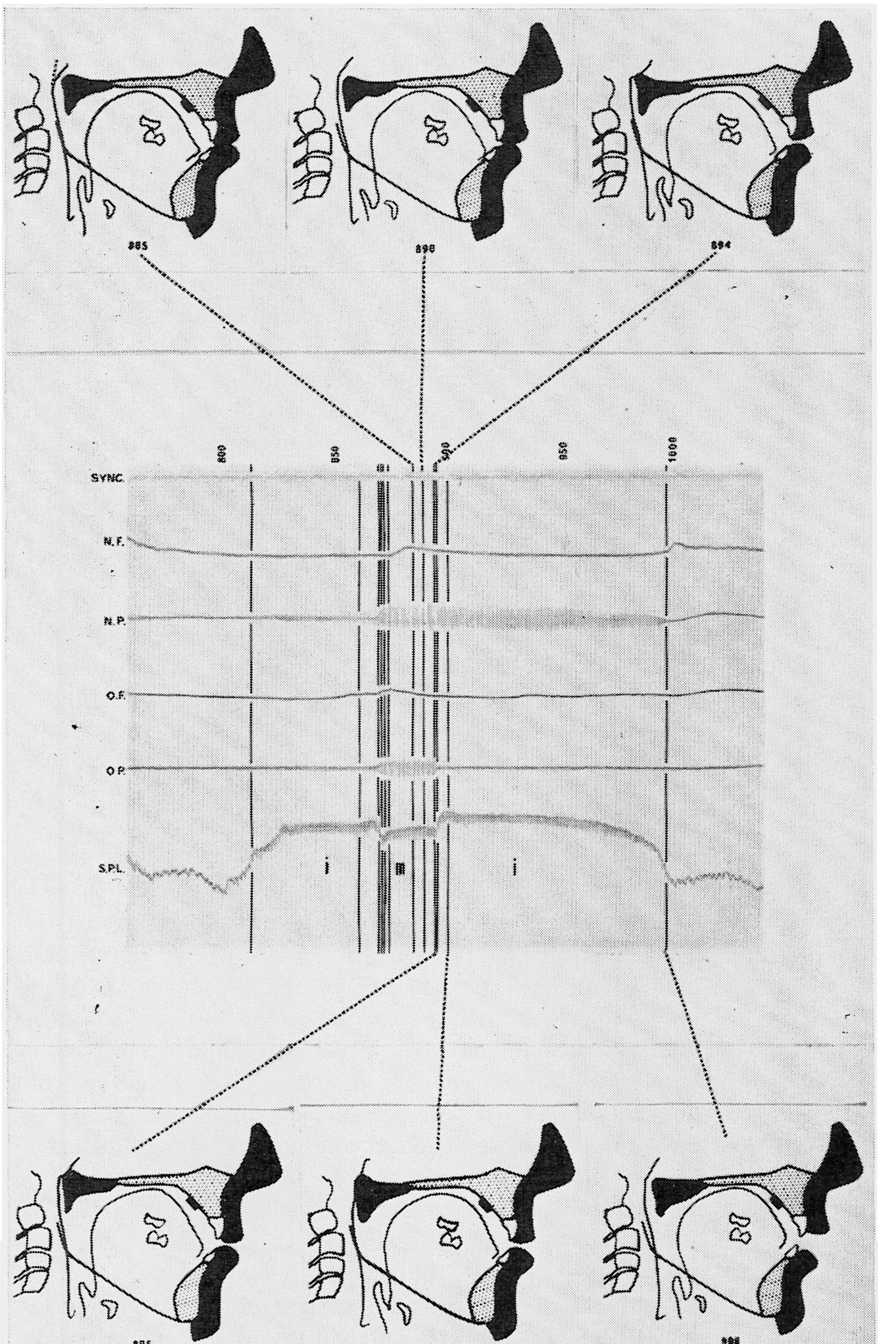


FIGURE 10. The tracings reproduced above display the temporal character of physiological events associated with the transition between /m/ and /i/ in /imi/ articulation.

recognized. These facts necessitate rigorous control and organization in study as well as a focus, restricted temporarily, to a limited number of phonemes.

The need to develop a broad multi-parametered reference for normal speech is apparent when the indivisible nature of the speech producing system is considered. In broad aspect, the source of speech power resides in the respiratory apparatus, which cannot be excluded for the sake of convenience in research. Respiratory as well as phonatory and articulatory functions are integral, inseparable components of the physiological system for speech generation. For this fundamental reason, broad design in research to develop significant parameters which can be related to respiratory and phonatory function is recommended.

The broad reference for normal speech is also needed to improve understanding of disordered speech, which is extremely complex. Although defective speech is known to result from faulty coordination, identification of deviations in patterns of movements may be exceptionally difficult. Defective sound generation may involve simple errors or many complexly interrelated deviations. For example, the direction and timing of movements may be relatively normal but the degree of movement restricted for some speech structures. Or the degree, direction, and range of movements may be essentially normal but faulty timing or ordering of movements may cause defective speech. Or, normal articulatory movements within the oral and pharyngeal tract, with proper timing of events, may exist but defective speech persists as a consequence of abnormal respiratory or phonatory activity.

Factors which can cause disordered speech are enumerated to emphasize the fact that disturbances in the timing of events as well as disturbances in phonatory, articulatory and respiratory functions may be responsible. The abnormal as well as normal speech apparatus is functionally integrated. To illustrate, respiratory patterns of cleft palate speakers with palatopharyngeal incompetence may be modified to generate: (a) excessive flow to raise intraoral pressure and overcome nasal escape; or (b) minimal flow to mask the audible feature of nasal emission. Similarly, abnormal laryngeal (glottal stopping) and nasal (grimacing) valving may be observed as the speaker attempts to manage the breath stream for intelligible speech sound production. These compensatory patterns exert recognizable modifications in the spectrum and in the aerodynamics of the breath stream as discussed by Warren and Ryon (*4*). To interpret the modifications recorded, multiple observations are needed.

The difficulty in interpreting one isolated parameter should perhaps be considered. Reduced nasal flow by itself may result from efficient palatopharyngeal valving but it might also result from reduced expiratory flow, weak voice production, nasal obstruction, constriction of the nares or glottal stopping. Thus, it is apparent that the interpretation of one parameter is facilitated by appraisal of others synchronously recorded. Since the fundamental purpose of study is to understand respective parameters as they relate to communication efficiency, observed phenomena should be related to the character of speech produced.

The instrumentation and procedures described here are currently being employed to study oral and palatopharyngeal valving in normal subjects and cleft palate patients before and after pharyngeal flap surgery. Because pressure and flow parameters are complexly inter-related and dependent upon physiological activity, independent and interdependent analysis of data is needed to supply basic information in areas of speech science and pathology.

**Summary**

This report describes instrumentation and procedures employed to obtain high speed cineradiographs (240 frames per second) synchronized with recordings of audio, intraoral, intranasal air pressure, and oral and nasal air flow. Analogue computer techniques providing additional parameters of sound power and volumetric information from sound and volume velocity recordings are outlined. Procedures in data accumulation and the type of data recorded are described and illustrated.

Data analysis is explained in accordance with the purposes of study: (a) to describe physiological events; (b) to specify the timing of events; and (c) to gain measures of pressure, flow, and sound power to assist in interpretation of observed events. The need to improve identification and understanding of normal and abnormal coordination as it pertains to communication is stressed. To fulfill this need, study of an increased number of parameters synchronously recorded is recommended, if not required. Implications of data, as described and illustrated, are discussed relative to research in areas of speech science and pathology.

reprints: *Dr. J. Daniel Subtelny, Chairman*
*Orthodontic Department*
*Eastman Dental Center*
*800 Main Street East*
*Rochester, New York 14603*

*Acknowledgements:* The authors express appreciation to the consultant on this program of research—Gordon E. Peterson, Director of Speech Communications Laboratory, Inc., Santa Barbara, California. The assistance of Jorge C. Mestre and Louis J. Russo is also gratefully acknowledged.

**References**

1. Peterson, G. D., and McKinney, N. P., The measurement of speech power. *Phonetica, 7,* 65–84, 1961.
2. Subtelny, Joanne D., Kho, G. H., and Subtelny, J. D., Multidimensional analysis of palatopharyngeal and oral valving during normal speech. Program abstr., 25th Annual Meeting of The American Cleft Palate Association, 1967.

3. Subtelny, Joanne D., Worth, J. H., and Sakuda, M., Intraoral pressure and rate of flow during speech. *J. speech hearing Res., 9,* 498–518, 1966.
4. Warren, D. W., and Ryon, W. E., Oral port constriction, nasal resistance, and respiratory aspects of cleft palate speech: An analog study. *Cleft Palate J., 4,* 38–46, 1967.
5. Worth, J. H., and Runyon, J. C., Integrating flowmeter for measuring unimpaired oral and nasal airflow. Submitted for publication *IEEE Trans. Bio.-Med. Engineering.*