Method of micro-analysis of sound films of behavior

WILLIAM S. CONDON, WESTERN PSYCHIATRIC INSTITUTE AND CLINIC, UNIVERSITY OF PITTSBURGH, Pittsburgh, Pennsylvania 15213

Initial viewings of sound films of human behavior resulted in the problem of how to divide the seemingly continuous flow of that behavior. This led to the development of methods and instrumentation that would permit variable-contrast scanning of the processes in order to detect points and criteria of segmentation. Since both speech and body motion are to be studied, instruments that would permit the analysis of each at a commensurate level were required. The instruments are primarily combinations of commercially available equipment that is relatively inexpensive yet reliable. These instruments, their use, and some of the preliminary observations are presented.

Initial confrontation with a sound film of human behavior often leaves the observer overwhelmed by a seemingly continuous flow of events. After repeated viewing, seeking for units and their forms of organization within this flux of events, it was felt that a central problem involved the development of a method and criteria for segmenting behavior. The method that evolved is essentially similar to that of the ethologist who observes animal behavior for many hours, seeking for patterns and regularities. In the present case, films of human behavior were viewed over and over until regularities began to be seen. This has, with modification, remained the primary approach. Instrumentation was concomitantly developed to permit more precise analyses and correlations.

In this paper, I will describe the method and instrumentation techniques developed in the microsegmentation of sound films of human behavior. Since the observer deals with both sound (speech) and body motion, the segmentation of both aspects will be discussed. The taking and processing of sound films will not, however, be dealt with. An excellent article on the taking and processing of research films is discussed elsewhere by the research cinematographer Jacques Van Vlack (1966).

SOUND SEGMENTATION

The problem of segmenting sound is a difficult one. It must be accomplished with some degree of precision and the length or duration of the various sound segments

must be related to body motion occurrences as accurately as possible.

Sound segmentation is relatively accurately achieved by using a film editor coupled with a sound reader. We use a modified 16-mm visual editor that projects a 9 x 7 in. image on a screen mounted 14 in. behind the viewer. This provides a somewhat larger image than the usual editor. A standard editor should be adequate, however. The sound reader is located adjacent to the viewer on the right at the appropriate distance. This is usually 26 frames, including the frame under the visual aperture and the frame under the audio head. (Most films are customarily sound synched at 26 frames ahead of the visual image to compensate for the necessary physical distance between the projected frame and the sound head on standard sound projectors). The film is manually transported through the viewer and the sound reader. Since standard film winders were somewhat unwieldy, special reel holders with knobs were constructed to hold and permit smooth transport of the film. This equipment will be referred to as an audio-visual analyzer, AVA for brevity.

With this equipment it has been possible to segment speech down to 1/48 sec relatively accurately. Each frame of the film is individually numbered, serving as a coordinating point for later correlating speech and body motion. Initially, the speech segments are related to the frame numbers. Later, using a better projector, the body-motion changes are related to the same frame numbers. This permits a comparison of speech and body motion at a commensurate level.

A usual procedure involves a preliminary reviewing of the film several times on a standard 16-mm sound projector. A transcript obtained from a ¼-in. tape of the sound on the film to be studied is necessary; linguistic training is a mandatory requirement for segmentation below the word level. The following example from a 48-frame/sec (f.p.s.) sound film of two psychiatrists talking spontaneously will seek to illustrate the process of segmenting speech. The phrase "ta ask you," taken from the total utterance "I was goin ta ask you why do you . . ." and spoken by one of the interactants, will be analyzed. Slowly transporting the film through the AVA and watching the viewer, one reaches the point of onset of the word "ta," which is slurred and articulated as "tuh." Spontaneous speech contains many such slurred elements. At the point of onset of

the "tuh," one can hear the /ə/ (a mid-central vowel with slight /t/ onset quality) and can see the frame number at which it occurs. This occurs at Frame No. 63. The investigator then continues transporting the film forward (with a back-and-forth motion) until the next sound change is detected. There is some sound distortion because of the manual operation that varies the speed. With some practice, however, one learns to transport the film at approximately normal speed and can detect points of sound change and the frame numbers at which they occur. Later, when analyzing the film with a better projector, the investigator can often corroborate this segmentation by the lip changes; /P/ and /b/ should occur when the lips are together, etc. The next sound change in the present illustration is /æ/ at the onset of "ask," and it occurs at Frame No. 66. The preceding scanning process is continued with the next change being /ε/, which begins at Frame 69. This procedure results in the transcript presented on the phonetics line of Fig. 1.

This provides the initial verbal transcription. One can also locate word and higher-level unit boundaries (phrases, utterances, etc.) in the same fashion. It is thus useful for investigators who do not wish to do microanalyses. Animal sounds can be segmented in a similar fashion. The higher suprasegmental (pitch, stress, and juncture) and paralinguistic analyses are done using a tape recorder.

BODY-MOTION SEGMENTATION

After the speech analysis has been completed to the satisfaction of the investigator, the kinesic (body-motion) analysis is begun. There are a variety of projectors available for such analyses. Most of these are described in photographic supply catalogues. The Bell and Howell Time and Motion Study Projector (TMP), a manual- and/or motor-operated 16-mm silent projector, has been the preferred instrument for kinesic analysis at this laboratory. It permits a manual, contrasting scanning of body-motion changes rapidly and in varying sequence combinations in the search for regularities. One frame can be viewed in contrast to a succeeding or preceding frame, or two frames can be contrasted with the succeeding four frames, or three frames contrasted with the next three frames, etc. Electronically controlled time-motion study projectors do not at present provide this important variable-contrast scan capability.

For intensive work, it was found useful to modify the projector slightly. In its standard operating form, there is a

Fig. 1. Microanalysis of body motion accompanying "ta ask you."

tendency to produce a slight backlash when manually transporting the film back and forth. While intensively analyzing a small portion of the film, this backlash produces pressure on the sprocket holes of the film that tear, necessitating removal of the film and splicing. This consumes time and tends to ruin that segment of the film for further viewing because of the splices. A solution to the problem involved the development of a relatively "free-floating" transport system. The upper reel arm of the projector is not used. Instead, a separate reel holder was constructed that sits to the front and slightly to the left of the projector, out of the line of the projected image. This is designed so that the film reel moves freely but not so freely that the reel turns by itself, which would result in the film spilling to the floor. A feeder trough, with a lid to hold the film in place, was constructed to fit on the front of the projector, angled to the left toward the auxiliary reel holder to guide the film into the projector. This eliminates pressure on the film when going forward. If the lever on the upper reel feed arm of the projector is placed in "rewind" position, the take-up reel on the lower reel arm will move freely, eliminating backlash when transporting the film backward. These modifications provide a free-floating transport system for the intensive back-and-forth analysis required in segmenting body motion. This is most useful when moving progressively through a film analyzing 10 or so frames at a time. It has resulted in considerable saving in film damage and in time. It is also important to have an auxiliary fan blowing on the gating mechanism of the projector to prevent excessive heating from continued use. A wide-angle projector lens (f2.0) provides a large image approximately 5 ft from the observer. The white matte projection screen provides much more definition than either a glass beaded or lenticular screen.

There are varying viewpoints concerning the method of transcribing body motion. These are not contradictory, but complementary and depend to a great extent upon the level and analytic perspective of the investigator. Birdwhistell (1952) and Schefflen (1965) have been pioneering figures in the communicational analysis of human body motion. A forthcoming book, *The Natural History of an Interview*, edited by Norman A. McQuown, will include a detailed workbook that should be of value to students and others wishing to analyze

FRAME NO.	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	
PHONETICS	---	---	---	↑	∂	∂	∂	∂	∂	∂	∂	∂	∂	∂	∂	∂	∂	∂	∂	∂	∂	∂	∂	∂	---
HEAD	U	U	Q,R,L	Q,R,L	U,S	(L,Q,R)	S	S	S	S	L,F,Q,R	L,Q,R	L,F,D	L,F,D	L,F,D	L,F,D	L,F,D	L,F,D	L,F,D	L,F,D	L,F,D	L,F,D	L,F,D	L,F,D	F(D,R)
EYES	Os	O,U	H								L	Lf	(L,U)	S	H										Cs
MOUTH	Os	O,W	Cs								(O,F)	S	OVS	Cs	(C,F)	S	F,Cs								O
R. SH.	RUS	RIVS	RIS										R,AD	S	RD,AD										H
R. EL.	Fs	Ss	Fs				FVS	Ps			P	Ps,EVS	(R,E)	S	E,PVS	Ef,P									Ef
R. WR.	ABs	H	Fs				F	Fs			Fs	FVS	Fs												Fs
R. FINGERS	1	H	Fs	F	Ff	Ff,AD	Fs	H																	E
	2	H	E	F	Ff	Ff,AD	H																		F
	3	Es	Fs	F	Fs																				F
	4	Es	Fs	F	Fs																				F
	T	Es	E				E,AD	E			Es	H													E

sound films. The present paper will focus on the analysis of the "flow" aspect of body motion rather than the relatively static postures. Intensive study of films led to the selection of points of change (of direction or speed) of the body parts to describe the behavior. These points of change are contrastively derived, and this will be illustrated below.

The body parts, with the emphasis on joints, are listed vertically at the left-hand edge of the transcription paper. Horizontally and vertically lined 17x11 in. data pads are used. For convenience and systematic technique, the listing begins with the head and works down: head, eyes, brows, mouth, right shoulder, elbow, wrist, and fingers (separately), left shoulder, elbow, wrist, and fingers, right hip, knee, and ankle, left hip, knee, and ankle, and finally the trunk. This list can be increased or decreased depending upon the focus of the investigator. Change of direction is primarily related to and transcribed in terms of the types of movement made by the joints: adduction/abduction, pronation/supination, flexion/extension, and the various rotational movements. These characteristic joint movements can be learned from a good kinesiology manual (Rasch & Burke, 1963). The descriptive symbology is largely composed of the first letter of the movement: e.g., F means flexion or forward, depending upon the body part; AB means abduction; P means pronation; Ps means slight pronation; S means supination, etc. Head, eye, mouth, brow, etc., movements are difficult to describe and some experience is required. Descriptive schemas for these types of movements are slowly being compiled.

Using the phrase "ta ask you," previously chosen to illustrate speech segmentation, the observer begins by

scanning the body motion activity across "ta" ("tuh"). It has been found useful to gain some familiarity with the grosser changes before beginning the more fine-grained analysis. For the actual microanalysis, one begins with the head and works downward. Figure 1 is from an actual microtranscription of body motion.

During the three frames (60, 61, and 62) preceding the onset of the word "ta," the head had been moving directly upward. This is determined by sweeping the TMP back and forth across these three frames, stopping on Frame 62. Continuing sweeping across these same three frames but now including Frame 63 in the sweep, a sudden change is detected in the head movement. This indicates a change in direction of movement at Frame 63. Now, beginning with Frame 63 (with the projector resting on Frame 62) and sweeping through Frames 63 and 64, no change is detected in the direction or speed of head movement. Then, going back to Frame 62 and sweeping through Frames 63, 64, and 65, still no change of direction is detected. However, when Frame 66 is included in the next sweep, a change is detected, indicating a change boundary. In order to describe the movement of the head more accurately, it is useful to sweep across Frames 63, 64, and 65 several times. It is also often helpful to contrast these frames as a unit with the preceding three frames and several following frames for a more accurate description of the direction of change. Thus, one can scan Frames 60, 61, and 62 as a continuous series, then Frames 63, 64, and 65 as a continuous series, etc. It is, again, this ability for variable-sequence scanning that makes the TMP more versatile than other types of projectors. Across the word "ta," the head sustains the complex direction of up, incline right, and left very slightly in

contrast to the upward direction of the three preceding frames and the incline right, left, and up slightly of the following three frames. These changes and the points of change are relatively clearly detectable after some experience. All detectably moving body parts are systematically described in the same fashion as the head.

As a result of this method of analysis, the body was observed to move in configurations of change that seem to be isomorphic with the articulated segmentation of speech. The word "ask," for example, was segmentable into four aspects: /ææææ/, /ÉÉ/, /sss/, and /kk/. The body also seemed to move in four configurations of change that were isomorphic with these four speech aspects. The body parts are analyzed each in turn, providing multiple criteria for the determination of the configuration boundaries. On-going modifications of change in speed and direction of the body parts tend to occur at the same point, to be sustained together for a brief interval, resulting in serially sustained configurations of change that appear isomorphic with the elements of speech. This is seen in the tendency for vertical alignment of the arrow points in Fig. 1, which are co-occurring points of change of the body parts.

Further contrastive analysis across wider dimensions suggests that the isomorphism between speech and body motion may occur across multiple organizational levels. Thus the body seems to regularly and simultaneously sustain movements of a body part or parts across syllable, word, phrase, and utterance length emissions. There is some suggestion that this may also occur in various ways across yet wider dimensions. What seems to be occurring is that the body parts sustain (in both direction and speed) different forms of configurations of change across the wider syntactic forms. These differences are detected by a contrast of the forms of these changes at the various levels. The word "ask" can be distinguished into the two aspects /ææææÉÉ/ and /ssskk/ in terms of its manner of articulation. If one contrasts the body motion across the first aspect in relation to the second aspect by sweeping across Frames 66-70 and then across 71-75, the body now appears to be describable as having two configurations. This contrast, however, is from a different perspective than the contrasts that resulted in the more micro four-part segmentation. The more micro changes tend to smooth out and are not attended to during the higher-level contrast. The following descriptive examples may help clarify the nature of the criteria for higher-level segmentation.

Across /ææææÉÉ/, the mouth begins to close and closes relatively rapidly, the right shoulder begins to rotate inwardly more rapidly with the elbow flexing slightly, the fingers begin to flex relatively rapidly as the finger-pointing gesture begins to change to a closed-hand formation. The head now begins a more marked turn to the left while inclining to the right. Across /ssskk/, the mouth opens slightly and the lips come forward slightly, the eyes now move left, the head moves left more rapidly while the inclination to the right decreases. The right elbow now begins to pronate rapidly and the fingers flex very rapidly, completing the closed-hand formation as the end of the word is reached. There is relatively greater rapidity of movement across the second aspect. The body parts appear to change velocity of movement in complex yet ordered ways in relation to the changing pitch and stress patterns of the speech. Kinesic pulsing tends to parallel the pulse form of the speech. The body will often move relatively slowly across the initial consonant, accelerate across the vowel, then slow across the terminal consonant in, for example, a consonant-vowel-consonant form. This will vary, however, depending upon the stress pattern of the total utterance.

The word "ask" also occurs as a total lexical unity. In contrasting the body motion across "ta" with that across "ask" and across "you," forms of change occur that seem to sustain across the emission of "ask" in contrast to the preceding and following words. Across "ta," the head moves primarily up, with a slight inclination to the right and very slightly left. Across "ask," the head moves primarily left, while across "you," the head moves left, down, and forward. During "ask," the arm moves rapidly and horizontally left across the front of the body in contrast to "ta," where it moves left very slightly and across "you," where it moves left and down.

Similarly, at a still wider dimension, the total phrase "ta ask you," is accompanied by the body in a form that contrasts with the preceding and following phrases. The total utterance again was "I was going ta ask you why do you . . .," with a pause following the terminal "you." Across "I was goin," the head moves up, the finger point forms and holds, while the arm moves out and up. Across "ta ask you," the arm moves left across the body, the head moves left with minor variation, and the finger point changes to a close-hand formation and holds. Across "why do you," the head moves to the right rapidly, the hand opens, the mouth opens rapidly, and the right arm now moves rapidly straight down to the speaker's lap. Thus,

the body may move in hierarchically organized patterns that seem to be isomorphic with the emergent syntactic organization. This observation has been tentatively called self-synchrony. (The hypothesis of self-synchrony is defined as an on-going isomorphism between the duration of articulated phonetic segments and body-movement configurations. Thus, the body will exhibit a configuration of change, which usually involves several body parts moving in a variety of directions but which are sustained together for a limited duration. The sustaining of these configurations parallels the emission of the segments of speech. The emphasis here is upon the term configuration.)

These "levels" of body-motion configurations emerge as a relation of the patterns of change of the body parts to the particular contrast perspectives that permitted them to become detectable. "Contrast," as described earlier, refers to unitary forms of change of the body elements as detected in the field of preceding, circumscribing, and following forms. The segmentation of the body motion across the word "ask" into four aspects, then into two aspects, and finally as a unitary form is not contradictory. These diverse, yet integrated, patterns of change are detectable "in" the behavior. The four-part segmentation resulted from contrast perspectives that differed from the contrasts that led to a two-part segmentation, which, in turn, differed from that which led to describing the body motion across "ask" as having a unity of form differing from that across "ta" and "you." These differing "levels" of forms of change emerge in terms of the diverse relationships of change and change rates the body parts sustain to each other over time. The difference between the body-motion patterns across /ææææ/ and /ÉÉ/, which defines that across /ææææ/ as having a distinguishable unitary form, is derived from different relational sets of changes of the body parts than the contrast between /ææææÉÉ/ and /ssskk/, which revealed /ææææÉÉ/ as a unitary form at a higher level. These emergent levels of organized changes of the body parts to each other appear to parallel, as indicated, the emergent levels of syntactic organization. The methodological-instrumental approach discussed above may permit an approach to the analysis of the relation of the organizational levels of speech and body motion to each other over time and their role in communication.

The relatively microscopic focus herein presented was not a matter of preference. Early work had focused on the relation of body motion, particularly gestures, to verbal meaning or meaning classes in the

speech. While a relationship often seemed to exist between some gestures and given meaning classes, the precise boundaries of the gesture and the meaning class were difficult to determine with existing equipment. There were other, more serious, difficulties. For example, if a gesture occurred with a smile and later a similar gesture occurred with a frown, did this alter the meaning of the gesture? Again, a psychiatrist's use of "um hum" in response to a patient seemed to have many different interpretations, depending upon the paralinguistic and body motion accompanying it. This resulted in some hesitation in abstracting the "um hum" apart from its form of expression. One can say the same phrase, using different pitch and stress patterns and/or accompanying body motion and markedly change the connotation. Occurrences such as these suggested that the total expressive manifestation of the speech and gesture be investigated. This led to a gradual focus on the more microscopic aspects of the behavior. The observation of self-synchrony emerged from these analyses and did not, initially, constitute a specific focus. The search for the psychological relevance of the patterned relationships of speech and body motion during behavior has not been abandoned. There appear to be basic regularities at many levels in behavior that may be useful in analyzing the organization of behaviors expressive of emotion, attitude, etc.

The body-movement patterns of the listener in the interactional situation were studied in the same fashion as those of the speaker. The listener was discovered to move in configurations of change that were isomorphic with those of the speaker, although the same body parts might not be moving or moving in the same directions as those of the speaker. Interactants participate in shared, dance-like movements during communication. This phenomenon has been called interactional-synchrony and is also often hierarchic in form. A speaker and listener may each sustain a pattern of change together across a phrase, then change together to sustain a different pattern across the next ensuing phrase, etc.

Several modifications of behavior were introduced to check the hypotheses of self- and interactional synchrony. (a) Two Ss on camera were "pulled out" of their interaction to each interact with a separate person off camera. No interactional synchrony could be detected between the two Ss on camera. (b) Delayed speech was employed to see if it disrupted the interaction. The voice of the speaker was delayed to the listener who wore ear

phones. The body-motion patterns between speaker and listener appeared to be nonsynchronous. When the speaker was subjected to delayed auditory feedback, he stuttered and became mildly self-dyssynchronous. This is discussed below. (c) The hypotheses of self- and interactional synchrony imply, for example, that the eye blinks of a speaker will occur synchronously with the change points of the rest of his behavior and also that those of the listener will be synchronously related to the speech/body motion patterns of the speaker. Over 1,000 eye blinks of 20 speakers and listeners were studied frame by frame. The speakers' eye blinks occurred synchronously with the segmental patterns of their behavior. The listeners' blinks occurred synchronously with the patterns of change of the speaker's behavior. These observations were felt to lend support to the hypotheses. Kendon's analyses also support the hypotheses of self- and interactional synchrony (Kendon, 1969a, b).

This approach has also been used to investigate a variety of forms of pathological behavior. Films of patients with aphasia, Parkinsonism, petit mal, stuttering, schizophrenia, and childhood autism have been studied. Many were found to display patterns of change that differed in subtle ways from the harmony characteristic of normal behavior.

Multiple high-speed cameras have been used to record a variety of subject matters that are then examined using the above methodology. The cameras are synchronized together, with one camera on the Ss and the other on the monitoring equipment. For example, two synchronized high-speed cameras (94 f.p.s. with sound), one on the Ss and the other on an EEG machine used as a polygraph, were used to examine behavior in relation to the polygraph pen patterns. The film of the polygraph pen movements was studied in the same scanning fashion as the body motion and related to the same frames that recorded the body motion. Many diverse correlational studies are possible using this multiple camera technique.

In another study, six "normal" Ss were filmed (each in turn) at 48 f.p.s., speaking normally and then under delayed auditory feedback. A microanalysis of the organization of change during both conditions revealed that under the delayed feedback condition all six Ss displayed a self-dyssynchrony that resembled a mild aphasia in contrast to the normal condition which was self-synchronous (Pacoe, 1968). This study exemplified to some extent the integrated use of the natural history and experimental methods.

Films of animal behavior were also studied at both micro and macro levels (Condon & Ogston, 1967). The pattern of behavior of various species and the regularities of form within those patterns may be amenable to detailed study using the above instrumentation and segmental procedures. For example, there may be shared movement patterns occurring in animal interactions that may be useful in studying the structure of the animal group. Also, nonverbal behaviors accompanying various vocalizations may aid in assessing communicational variations in those vocalizations. In essence, it might prove helpful to study the total animal behavior, not separating vocalization and body motion into isolated aspects.

CONCLUDING REMARKS

Methods and instrumentation developed in the study of sound films of human behavior have been presented, including some of the observations. A surprising degree of order was found to pervade that behavior, particularly in the relation of speech and body motion.

This approach may also be feasible for a wide variety of studies where the processes are amenable to storage on sound films. These films can then be re-viewed in variable-scan sequences as often as desired in the search for the regularities in those processes.

REFERENCES

- BIRDWHISTELL, R. L. *Introduction to kinesics*. Foreign Service Institute, U.S. Department of State, Washington, D.C., 1952.
- CONDON, W. S., & OGSTON, W. D. A method of studying animal behavior. *Journal of Auditory Research*, 7, 359-365, 1967.
- KENDON, A. Movement coordination in social interaction. *Acta Psychologica*, in press.
- KENDON, A. Some relationships between body motion and speech: An analysis of an example. In A. Siegman (Ed.), *Research in interview behavior*. New York: Pergamon Press, in press.
- MCQUOWN, N. A. *The natural history of an interview*. New York: Grune & Stratton, in press.
- PACOE, L. V. Micro-segmentation of sound films: Reliability and control. Paper read at the American Psychological Association meetings, San Francisco, August 30, 1968.
- RASCH, P. J., & BURKE, R. K. *Kinesiology and applied anatomy*. (2nd ed.) Philadelphia: Lea & Febiger, 1963.
- SCHEFLEN, A. E. Stream and structure of communicational behavior. *Behavioral Studies Monograph No. 1*, Eastern Pennsylvania Psychiatric Institute, Commonwealth of Pennsylvania, 1965.
- VAN VLACK, J. Filming psychotherapy from the viewpoint of a research cinematographer. In L. A. Gottschalk and A. H. Auerbach (Eds.), *Methods of research in psychotherapy*. New York: Appleton-Century-Crofts, 1966. Pp. 15-24.