THE CONTRIBUTION OF KINESIC CUES TOWARD SPEECH COMPREHENSION

by

WILLIAM ROGERS

A dissertation submitted to the Graduate Faculty in Speech and Hearing Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, the City University of New York.

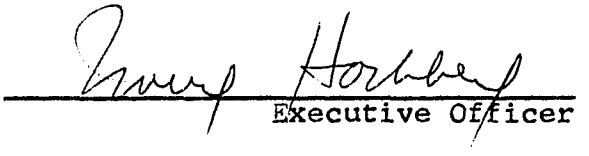
1977

This manuscript has been read and accepted by the Graduate Faculty in Speech and Hearing Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

2-1-78
date


Chairman of Examining Committee

2-10-78
date


Executive Officer

Professor Louis Gerstman

Dean Norman S. Rees

Professor Stephen Thayer
Supervisory Committee

The City University of New York

Abstract

THE CONTRIBUTION OF KINESIC CUES TOWARD SPEECH COMPREHENSION

by

WILLIAM ROGERS

Advisor: Professor Louis Gerstman

Two experiments were conducted to assess the extent to which a speaker's visible body movements can improve verbal comprehension for the listener. Listeners responded to multiple-choice items designed to test their comprehension of 12 videotaped spoken utterances which had been obtained by asking speakers to describe objects in motion (e.g., a tennis ball, a car, spraying water). Sixty subjects each responded to stimuli in one of three presentation conditions (audiovisual, audiovisual without lip and facial cues, and audio-alone) over four signal to noise ratios. The results indicated that visual cues can at times significantly improve comprehension scores, even with lip and facial cues not present. A number of possible variables which may be related to kinesic utility for speech processing are considered and a theoretical model of aural-visual speech processing is discussed in relation to future study and applications.

Acknowledgements

I would like to strongly express my gratitude to Dean Norma S. Rees for providing me with needed encouragement, advice and support for my project, especially at the time of its inception. Without her support I might not have succeeded in pursuing it. Dr. Louis Gerstman, my mentor, expertly provided just the proper balance of direction and freedom in overseeing my general doctoral studies and in overseeing the progress of my disseration activities. Without his help, my research design and data collection and analysis would have suffered greatly.

I would especially like to thank the several hundred students of mine who participated in my experiments, earnestly, many times offering valuable comments from the subject's perspective. More than once, their unqualified cooperation gave me the energy to continue in the face of discouraging research difficulties.

Two individuals were responsible for my developing the needed enthusiasm to devote five years of study to the topic of nonverbal communication: Dr. Stanley Jones, who was my teacher and later my colleague, and Dr. Ray Birdwhistell, whose wonderful writings on the topic captured my attention and motivated me to study the aspect I chose.

Finally, I would like to express fond thanks to a fellow student and colleague, Ann Jablon, who, from the

very beginning to the very end, continually listened to, and constructively commented on, my frequent worries, concerns and reports about my research activities.

William Rogers

January 19, 1978

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

INTRODUCTION

In recent years, interest in studying the role played by nonverbal behavior within human communication has steadily grown beyond the limited set of studies carried out in the earlier part of the twentieth century. The earliest studies on the topic (cf. Darwin, 1872; Dunlap, 1927; Schlosberg, 1941) were concerned with describing the use of facial expressions to convey emotional meanings. The question ultimately arose in the area about whether such behaviors were biologically determined (and, thus, universal) or culturally learned. After many years of research controversy, it now seems reasonable to say that both the biological and cultural explanations are needed to account for the often subtle and considerably diverse use of facial expressions by members of the world's various language groups.

With the development of clinical psychology such behaviors as postural shifts, eye movements, and hand positionings were studied by the clinician (cf. Freud, 1901) as a clinical technique for "reading" the unverbalized feelings of the client. Although more often than not these studies were anecdotal, non-systematic, and unadvisedly dependent on the judgement of one individual, later systematic investigations (cf. Ekman, 1964;

Mehrabian, 1969; Exline, 1966) did at least bear out that certain human feelings concerning interpersonal power, affect, and involvement could be reliably signaled, with limitations, by postural, eye and hand behaviors.[1]

Probably the earliest study of communicative body movements, or kinesic behavior, concerned what are commonly called gestures. In fact, one finds scholarly comment on the use of gestures as an alternate means of communication as far back as the writings of Plato (Hewes, 1976). Although gesticulation was the first type of nonverbal communication to receive formal attention, the expectations on the part of early workers in the area--that a grammar and dictionary of gesture could be developed--discouraged study here when this objective proved elusive (with some very minor exceptions, cf. Munari, 1963).

It was in the early work of Birdwhistell (1952) that the notion of <u>context</u> elucidated why grammars and dictionaries (in the linguistic sense) were probably not practical objectives, because a kinesic behavior (e.g., a head nod) could mean one thing (e.g., "I understand") in one situation (after a question) and mean something else (e.g., "I want that one") in a second situation.[2] In addition, the easily defined <u>gesture</u> was replaced by the <u>kinesic act</u>, which because it includes more types of movements assigns, then, a greater variety and richness to the domain of speech-related nonverbal behavior. With the above two conceptual clarifications the potential was created to

study speech related kinesic acts more profitably.

And recently, interest in studying body movements occurring during verbalization has been rekindled, probably because of the clarifications described above and because of the improved potential to dissect communicative body movements for study made possible with the current use of videotape and 16 mm sound film as a research tool (cf. Condon and Sanders, 1975). However, the more recent studies attempt to describe the function and structure of verbal -kinesic message units without imposing a grammatical model.

For example, Condon (1966, 1970, 1974) has reported that at least on the motor level of behavior a significant amount of synchrony exists: (1) between the speaker's body movements and the acoustic rhythm within his utterances (self synchrony); and (2) between the body and speech rhythms of two or more persons in the same spatial environment (interactional synchrony). Condon and Sanders (1975) further suggest that young infants will synchronize their gross body movements with the speech output (live or audio recorded) of an adult, but not to such non-speech sounds as tapping on a table. This tendency for a person to synchronize his own body movements with his speech rhythms seems to indicate that the two activities are probably coordinated at some internal behavior-planning level, possibly facilitating the moment-to-moment utterance processing taking place in both the

speaker and the listener.

On a different level, Dittman (1972) observed kinesic activity in relation to hesitations within utterance clauses. He reports that hesitations tend to occur in the early part of utterances, while movements of the body parts tend to occur at utterance start positions. He also noted that the high information word in the utterance receives the greatest stress (vocally and kinesically) and that head nods frequently punctuate boundaries between language segments having one main clause. Dittmann argues that the temporal relation he observed between body movements and speech hesitations implies that within-utterance body movements are motor manifestations of underlying cognitive activity controlling speech and is more manifest at times when speech does not flow evenly (or does not exhaust the underlying activity). Kendon (1972) draws this same inference, while he, in addition, goes on to describe how the speech flow and kinesic flow comprise a compatible pair of hierarchically structured (smaller units combining into larger ones) outputs which coincide at the prosodic segments level of communication. In his words, the situation is that of a "single utterance with two outputs."

A number of researchers have employed the strategy of examining slow motion videotaped or sound-filmed records of conversation samples, lasting as long as sixty minutes, over very extended periods of time (in several cases ten years or more) in much the same way as a biologist might

examine the structure of a substance using a microscope.

A prime example of this strategy is the seminal work of Birdwhistell who reports (1966) that certain semantic aspects of an utterance can be kinesically marked in predictable ways. These markers concern such things as personal pronouns, verb tenses, distal relations (here, there, these, those, etc.) and pluralization, in addition to reflecting action relations between subject and object. DeLong (1974) observed that children frequently end their utterances in child-to-child conversations with a leftward and downward head movement. And following a different line of thought, Lindenfeld (1971, 1974) reports that to a limited extend leg movements start and stop at phrase boundaries. Scheflen (1974) discovered after many years of observation that two communicants will reliably segment an entire conversation with the use of kinesic markers at several levels roughly equivalent to the segment-levels in writing of: sentence, paragraph, and essay. Finally Freedman (1973a, 1973b, 1974 devised a category system involving body-focused movements (nervous touching of hand to hand, hand to head, etc.) and object focused movements (hands making configurations in the air). The latter generally occur during speech activity.

To summarize what the above studies reviewed here indicate, several things can be said, specifically: (1) nonverbal behavior is synchronized with the rhythm of the utterances within which it occurs: (2) the extent of the kinesic "output" in utterances can vary qualitatively and

quantitatively, probably depending on underlying cognitive variables and on the type of verbal material being expressed; (3) kinesic material is temporally structured similarly to the verbal material it occurs with, although not with the mathematical precision of a grammar; (4) within utterance kinesic behavior is qualitatively different than kinesic behavior occuring in non-speech situations; and (5) kinesic behavior can be "about" the meaning of at least some aspects of the verbal material in the utterance.

A plausible description of the role played by kinesic behavior, especially in the communication of cognitive material, seems to emerge with kinesic behavior assisting the functioning of verbal processing to varying extents. More specifically, it seems that the within utterance kinesic flow is intimately involved with the speech flow neurologically, temporally, semantically, and structurally in a supportive way.

This supportive role (as opposed to a primary role) taken by kinesic behavior is consistent with the gestural-type language-origin theories asserted by some students of the phylogeny of language behavior (cf. Stokoe, 1975; Steklis, 1975). These theorists assert that gestural behavior now maintains a vestigal status, although it probably preceded verbal behavior, in time, as the principal form of human communication and had the potential to develop into a full-fledged symbolic system had sound

communication not proven more efficient (actually it did develop fully for the deaf who use sign language while it developed partially in such instances as the sign language used by American Indians (cf. Hewes, 1974).

In addition, it seems plausible to state that within speech body movement is, in part, learned, and thus culturally coded to some extent. For example, Freedman (1974) reports that speech related kinesic acts, as he describes them, were almost completely absent within a group of congenitally blind subjects which he had compared with a sighted control group. Also demonstrating the learned aspect of this type of kinesic behavior is a study by Michael and Willis (1968) which found that children having at least one year of school were significantly better at decoding common gestures than were children without at least one year of school, regardless of age.

On the other hand, several studies have also demonstrated that the language hemisphere is active in the production of speech related kinesic acts. For example, DeRenzi , Pieczuro, and Vignolo (1968) compared brain damaged persons and found that those persons who evidenced either ideational apraxia (a decreased ability to demonstrate the use of objects) or ideomotor apraxia (a decreased ability to perform an intransitive gesture such as a salute or a sign-of-the-cross) were very likely to also evidence aphasia (a decreased ability to use verbal symbols). They termed the individuals who were both

apraxics and aphasics as suffering from a conceptual disorder. The above evidence seems to indicate that gestural behavior which requires some specific plan of particular motions is processed by the dominant hemisphere.

Supporting the dominant hemisphere notion are several studies by Kimura (1973a, 1973b, 1974). For example, she reports that the hand gestures that occur during the speech of right handers were more likely to be performed by the right hand than by the left (3 to 1). However, those gestures occuring during non-speech activity were just as likely to be performed by either hand. Left handers are less clear in this distinction. Kimura and Archibald (1974) also report that persons with left himisphere brain damage find it difficult to copy visually presented but meaningless hand movements, while they show no impairment for performing flexions of the fingers or for copying static hand positions. Both sets of studies described here indicate that the behavior considered presently is at least in part genetically predispostioned.

The foregoing studies have indicated, indirectly, that nonverbal behavior ought to facilitate the decoding of spoken messages, for a number of reasons. However, it should be said that they did not attempt to measure <u>how</u> functional this support might be when processing various kinds of utterances. In fact few studies have made any attempt to investigate the crucial question of functionality.

One of the earlier of this small group of studies (Chapanis, 1972) failed to find support for a functional relationship. However, it should be conceded that the author was addressing a wide range of questions, this particular one being a small aspect of his text. Briefly, he asked pairs of individuals to cooperate in solving a problem while he manipulated the channels open for particular pairs (i.e., normal face to face, telephone line only, hand writing, and typewriting). Although face-to-face condition took slightly less time for subjects to complete the problem than did the telephone condition (kinesic cue deprived), the difference was barely worth mentioning. It might be pointed out however, that a ceiling effect may have caused the absence of a dramatic difference and that the type of verbal material (abstract) communicated was the least likely to benefit from gestural support.

The above speculations seem credible in light of two very recent and related studies. Graham and Argyle (1975) asked English and Italian speakers to communicate the identity of two dimensional shapes (some of which were called "high codable" because it was possible to easily find words to label aspects of the shape, as in "It is a distorted triangle with rounded edges;" and some were called low codable" because they were very irregular shapes without the access to convenient labels). The authors either permitted speakers to freely use their hands to

help communicate the identity of the figure or they did not permit the use of hands by requesting speakers to fold their arms while speaking. The decoders drew what they thought to be the actual two dimensional stimuli, based on hearing the speakers' messages. These drawings were then judged, systematically, as to how accurately they approximated the original drawings. As it turned out, the main effect (permitting the use of hand gestures or prohibiting them) was found to be highly significant for both English and Italian speakers ($p < .001$).

In addition, two trends were found: (1) Italian speakers utilized these facilitative hand movements more effectively than did English speakers ($p < .001$); and (2) low codable test items benefited more than high codable ones when hand movements were allowed. It should also be pointed out that the experimenters took very careful precautions to avoid threats to experimental validity during all phases of the study. They then performed an Ex Post Facto analysis to evaluate if and to what extent the requesting of speakers to fold their arms while speaking may have disrupted the execution of speech behavior. However, they found no differences across the two conditions with respect to a number of extraverbal measures designed to assess this possibility.

In a related study, Graham and Heywood (1975) investigated, using the same materials described above, what effects the elimination of gesture might have on such

things as semantic content of utterances and the total time spent speaking, etc. Their findings indicated that the inhibition of gesture in the communication of two dimensional forms resulted in: (1) quite expectedly, a greater number of words and phrases denoting spatial relations: (2) a fewer number of such demonstratives as "that," "here," etc. and (3) a higher proportion of time spent pausing during speaking time. In addition, the low codable and high codable items were found to differ significantly in their respective effects on the speakers' outputs over a number of extraverbal measures such as total number of words used, total number of pauses, and number of descriptive noun and verb phrases.

What the previous three studies indicate is that within utterance kinesic cues can play a functional role, although the importance of this role varies depending on the type of verbal material being expressed. What they also indicate is that a ceiling effect may operate in certain situations, where adding kinesic cues either does not improve listener performances (such as was found in the dyadic abstract problem-solving situation) or adds a significant but small improvement. What the previous three studies have not addressed is the measurement of what percent improvement in comprehension results from kinesic cues. Nor have they directly investigated the kinesic role in terms of actual comprehension since they observed time taken to problem solve or the degree to which a recreated drawing matched an original one.

No study seems to have investigated the kinesic role directly, as have investigators of the related but conceptually different role of lipreading in association with speech comprehension.

One early study in particular (Sumby and Pollack, 1954) demonatrates how the contribution of kinesic cues might be investigated. In the Sumby and Pollack study, a group of speakers was requested to say individual words which listeners were then asked to identify from specific word lists of varying lengths (8-256). The major manipulation in the study was whether listeners could see the speaker's face (speakers were asked to only move their lips) or not. In addition, six signal to noise ratios were employed by adding a white noise background to the speakers' voices, the intensities of both tracks being controlled electronically. The findings of the investigation were that: (1) lip cues did significantly add to the accuracy of word identification scores; (2) the size of the contribution of lip cues to improved accuracy was considerably greater (40-80% depending on word-list size) where the signal to noise ratio was low (-30 db) than when it was high (0-10% at 0 db); and (3) the size of the contribution was significantly greater where the task was easier (i.e., smaller word-list).

The authors also examined the operation of a ceiling effect they had found, by evaluating how much of a contribution lip cues rendered comprehension <u>relative</u> to the

room available for improvement over the audio-alone scores at the six S/N ratios. In comparing the observed improvement divided by the available room for improvement, they stated that:

> This ratio is approximately constant over a wide range of speech to noise ratios. Specifically, for the 8-word vocabulary the ratio increases from about .81 at a S/N ratio of -30 db to about .95 at a S/N ratio of -6 db. For the 32 word vocabulary the ratio increases from .77 to .81 over the same range.... Practically speaking, however, since there is a much greater opportunity for the visual contribution at low speech to noise ratios, its absolute contribution can be exploited most profitably under these conditions.

Although the above study was thought a good model for the development of a functional type speech-support kinesic study, several things about the study ought to be stressed: (1) it did not deal with natural running speech, as speakers were asked to say one word at a time; (2) speakers were not permitted to move kinesically; and (3) speakers were not attempting to verbally communicate in the usual sense since they were articulating items on a word list. However, the study does suggest the importance of the variables of noise and linguistic context in relation to carrying out multi-modal studies.

In general terms, what the present study sought to accomplish was to measure the contribution of kinesic behavior to the comprehension of speech. It also sought to describe how such variables as type of verbal content, noise and length of utterance might influence whatever visual contribution might be found. Specifically, it was

asked:

1. Can it be demonstrated that the accuracy of the receiver's comprehension will be higher when visual body cues (e.g., hands, face, shoulders) accompany verbal utterances than when such cues are absent?
2. Will noise have less of a detrimental effect on comprehension when kinesic material is present than when such cues are absent?
3. Will certain types of verbal content (e.g., verbs) describing action) have a greater comprehension score potential with visual cue assistance then will other types of verbal content (e.g., abstract nouns)?
4. Will any of the three above questions need to be qualified by such variables as sex of speaker, sex of listener, or length of utterance?

A casual observer might consider some of the above questions unnecessary and argue that it is obvious that kinesic material can only improve verbal comprehension. However, students of nonverbal behavior (cf. Haley, 1963) have often pointed out that visual cues accompanying an utterance can indeed support the verbal content, but can also contradict that content or carry information about some other aspect of the communication. Thus, the potential for accompanying visual cues to add nothing to the semantic content or even to contradict the semantic content exists in addition to the common sense notion that gestures assist

in the communication of ideas (it is tangentally interesting to note here that many people feel it is not good to "speak-with-your-hands" while engaging in speech).

PRELIMINARY EXPERIMENTS

With the foregoing objectives in mind, a pilot study (or more accurately a probe) was conceived and carried out, based on the concept of the speech act [3] as a focal point. The basic goal of the first pilot was to investigate how accurately receivers could detect a given speech act without hearing the verbal content. It was assumed that such information about whether an utterance was a question, command or assertion would be coded within the speakers kinesic performance, especially since such information is not frequently made explicit in the verbal content, but is rather implied by such things as voice intonation (which prior study had shown (kendon, 1972) was closely related to body movement).

The study was carried out by having individuals, in pairs, participate in a structured interview which was so programmed that participants would ask questions, make positive and negative assertions, and give directions. The performances of these speakers were videotaped with camera in full view of both participants. Thirty-eight utterances from the total generated by the 10 participants were selected as being examples of interrogative, assertive, negative, or directive speech acts. The presentational videotape was edited electronically in such a way that

stimulus types were randomly sequenced, with blank spaces between successive items on the tape. The set of utterances ranged in duration between three and nine seconds of tape time. Unfortunately, because utterances were frequently edited out of running speech, stimulus foreperiods were not more than one second each.

A response sheet was constructed within which each of the 38 stimuli was represented by one multiple choice question, with the same four possible answers for all items: (1) interrogative, (2) assertion, (3) negative, and (4) request. Respondents would view each of the videotaped utterances, in turn, and mark the answer sheet at the end of each presentation of an item. Within this design no control group was used since the hypothesis was that the respondents' accuracy scores would be significantly higher than what could be expected by chance (25% in this case).

Twenty four undergraduates (12 male and 12 female) were recruited from a basic course in communication at Queens College in New York and asked to view the tape and answer multiple choice questions. Subjects were tested in groups of two to three at a time. The results indicated that: (1) subjects did better than chance expectancies would predict at identifying questions and negatives but not at the remaining two types (assertions and requests): (2) although subjects did better than what chance expectancies could explain, the effect was not a strong one (the overall hit rate was 34%, 9% over what chance scores should be). Several

things ought to be considered, however, about the task required of respondents. First, the receivers were asked to view the materials out of their original contexts, which makes the task harder. Second, the sound, which normally accompanies speech perception, was absent, making the task somewhat unnatural for normal hearing subjects.

A study discussed earlier is probably significant in the evaluation of the results of the first pilot. Sumby and Pollack (1954) found that lip cues facilitated speech comprehension best when the task was an easy one. Since the task employed presently was a relatively difficult one, the usefulness of visual cues may have been at their lowest potential. Thus, it may not be so surprising that subjects did only slightly better than chance expectancies would predict. It is quite possible, however, that a different type of design would have generated more dramatic results.[4]

A second pilot study was carried out which attempted to evaluate the contribution of kinesic cues to speech comprehension more directly than had the first pilot. In this second investigation a control group was used to establish a base line with which to evaluate the kinesic contribution. That is, one group would both see and hear the stimuli while the other group would only hear. Also, the effects of noise were also included within the design as a major manipulation.

This second study was carried out by first constructing a stimulus tape consisting of brief excerpted utterances

broadcast on several daytime television soap opera programs. A set of 12 utterances was obtained in this manner. Each utterance was transcribed for later analysis. Twenty four observers were then asked to view each of the 12 items, in turn, and to write the content of utterances as accurately as possible after each individual presentation. Observers' transcriptions would then be compared with the correct transcriptions of the 12 items and scored for errors (i.e., missing words and morphemes). Six different observers responded to the 12 items in each of the following presentation conditions: (1) aural and visual cues available; (2) aural cues only; (3) aural and visual cues along with a white noise background; and (4) aural cues only with a white noise background.

The results indicated several things. First, adding visual cues decreased listeners' comprehension errors both with noise present and with noise absent. Second, adding noise increased errors at all times. And third, the improvement resulting from adding the visual material was significantly greater in the noise condition than in the no-noise condition. These findings were consistant with the Sumby and Pollack findings, thus providing encouragement for further investigation.

A third and final pilot study was undertaken which hopefully would test the actual design of the main experiments to be performed. This last pilot was carried out by recruiting seven undergraduates from a basic course in

communication at Queens College in New York and asking them to view a series of 8mm silent filmed excerpts of various types of moving objects (e.g., a car making a broken-u turn, a pigeon walking in a park, a mobile with hanging cardboard owls, etc.) which had been collected for use in the study. There were 15 such 8mm recordings of moving objects which each of the seven speakers viewed, in turn, and then attempted to describe to a second person who would draw a diagram of the 15 communicated-about actions. As each of the seven speakers went about describing the 15 actions seen on the silent 8mm film to a second person, the speaker was videotaped by an acknowledged but hidden camera and microphone. This procedure yielded 105 videotaped descriptions in all. With the use of electronic editing, a stimulus presentation tape was constructed consisting of 21 items.

The 21 items on the stimulus presentation tape were chosen so that all seven speakers would be equally represented and that each of the original 15 filmed actions would be described at least once, with six of them being described twice. The number of words in the utterances contained on the stimulus tape ranged from 11 to 57. Receivers were recruited from a second basic course at the above school. Twenty four in all viewed the 21 stimulus items and responded by writing what they thought they heard in each stimulus immediately after hearing each item. The experimenter would wait until each subject had finished writing before going on. Four conditions were used to manipu-

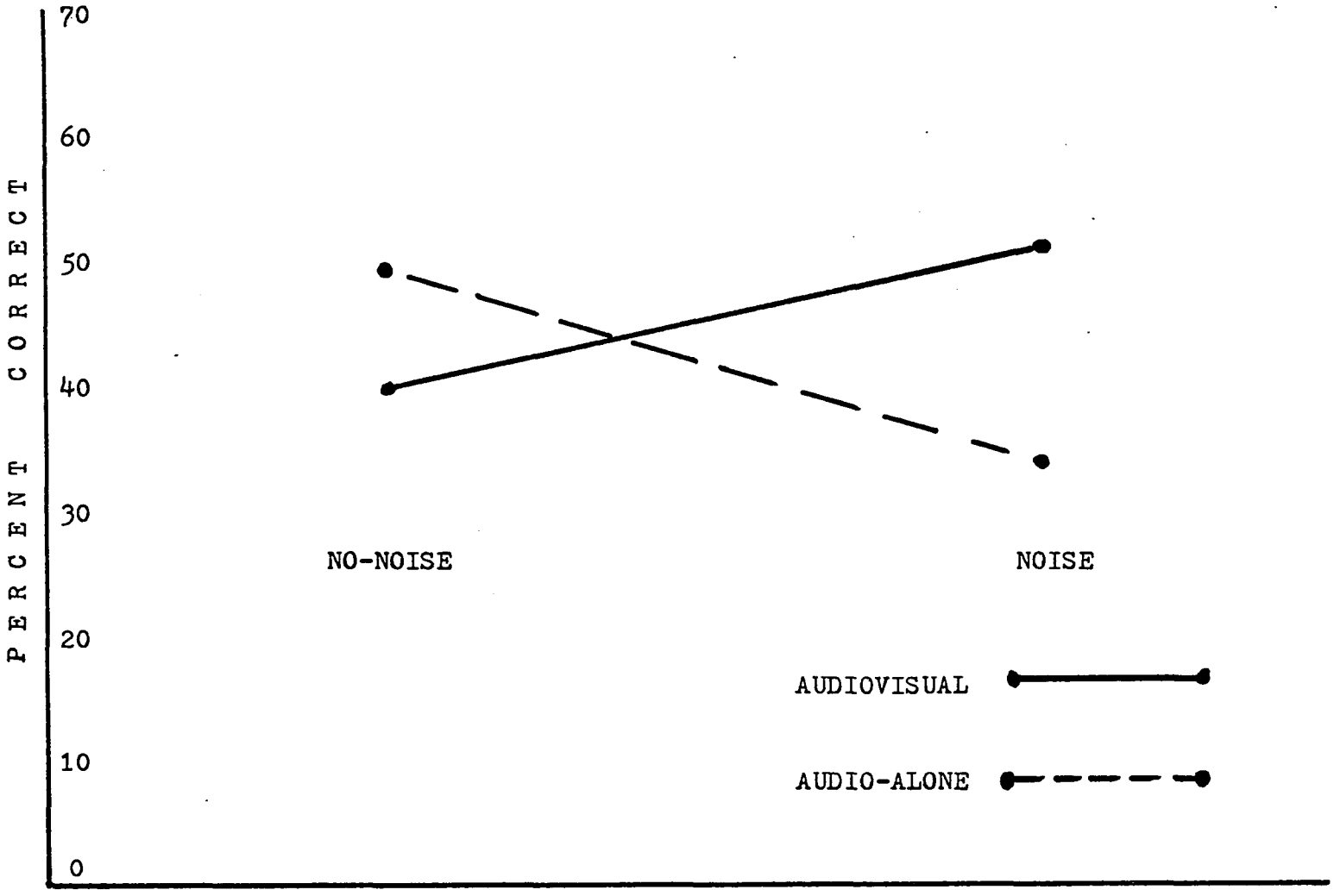
late the availability of visual cues and the use of white noise. Again, response sheets were scored for errors in terms of missing words and morphemes. Twenty four subjects were randomly assigned to one of the four conditions and tested within respective groups of six.

The results are shown in Figure 1, where it can be seen that some support for the main hypotheses was found, although to a limited extent. In the two noise conditions, the addition of visual cues shows the predicted effect, but in the two no-noise conditions the effect was not obtained. In addition, the variable of length-of-utterance appeared to be inversely related to the overall accuracy of scores, but not without exceptions. There were no significant sex differences for speakers or listeners, while some speakers appeared to be more successful communicators than others, with respect to listeners' accuracy scores.

Considered together, the three pilot studies suggested several things: (1) it is probably not wise to test the functional uses of the kinesic flow with respect to speech by removing the speech track from the task entirely, as the functional value of speech related kinesic cues probably only thrives with greater degrees of a vocal context; (2) the task given to the listeners should not be so difficult that it disrupts their abilities to relate body cues to some framework of linguistic meaning; (3) the type of verbal material being expressed by speakers ought to be selected within some particular semantic domain; (4) rather than

Figure Caption

Figure 1. Mean number of errors made by six subjects in each of the four conditions of Pilot 3.


70
60
50
40
30
20
10
0
PERCENT CORRECT
NO-NOISE
NOISE
AUDIOVISUAL
AUDIO-ALONE

simply using four conditions, two with noise added and two without noise, the signal to noise ratio of stimuli ought to be used as a major parameter involving three or possibly four different signal to noise ratios to be used in both the aural-only and aural-visual conditions; (5) the comprehension test used in the second and third pilots, where respondents wrote down what they thought they heard, should be replaced by a multiple choice format because the original test was very fatiguing for subjects (many made this complaint) and because it may have introduced an unwanted subject encoding-skills factor into the data (that is, subjects had to remember material and then recreate messages).

EXPERIMENT 1

The purpose of Experiment 1 was to formally measure the functional support provided by such visual cues as head nods, eye movements, lip-facial movements, hand gestures, and arm and shoulder movements to the accuracy of speech comprehension. Furthermore, the experiment sought to describe how signal to noise ratio and semantic type of utterance would further influence this functional support. It was predicted that the accuracy of scores would be greater when visual cues were available, especially at low signal to noise ratios. It was also suspected that certain types of semantic content would benefit more from the inclusion of visual cues than would other types although there was too little previous evidence to be more specific other than that some categories would achieve significantly greater support from visual

cues than others. The following hypotheses were generated:

1. Comprehension scores of listeners will be higher when they can see[5] the speaker than when they cannot.
2. The advantage of being able to see the speaker will be greater when signal to noise ratio is low than when it is high.
3. The types of semantic information tested for in the study will benefit, differentially, from the advantage of listeners being able to see the speakers.

Method

Audio visual materials. The presentational stimuli for Experiment 1 were patterned after those used in the last pilot study. However, it was the experimenter's judgement that speaker samples of a higher technical quality were needed and so a new set was produced. Prior to this, however, a re-edited 8mm film consisting of various novel movements was constructed similar to the film used in the third pilot but with a more diverse set of elements. The following brief scenes were assembled each lasting about ten seconds: a car making a series of turns, a truck entering a parkway, a pigeon walking in a park, two cardboard owls dangling on a mobile, an ornamental fountain spraying water, a tennis ball bouncing into a corner, a vehicle called a tram docking into an elevated outdoor station, and a movie projector performing its operations.

Speakers. Eight speakers (four female, four male) were recruited from a basic course in communication at Seton Hall

University and asked to perform as "actors." Taken in pairs each person would view one of the items contained on the 8mm novel movements film and attempt to describe what happened in the brief action to the second party, who would draw a diagram of the action based on the verbal message. Speakers were given practice performances on items in the 8mm film which were not to be used for that particular individual, in order to familiarize each person with the procedure and to give them time to adjust to the situation. Relaxing the speakers was considered very important since the experimenter's experience with the pilot studies was that nervous speakers tend to diminish their kinesic movements.

The speech performances (enacted about the 8mm film items) to be actually used in the experiment were recorded on videotape, with sound, using a small camera placed in view but discreetly positioned between two cluttered tables in order to diminish its intrusiveness. Practice items were not recorded, although it was not possible for participants to tell when the camera was recording from when it was not. All eight speakers participated in the same way, while half of them, in addition, also made short statements about abstract topics (such as on the purpose of a monetary system in a society), which were also recorded. The above procedure generated a presentation tape consisting of 12 items with several seconds of blank space foreperiod time placed just prior to each item.

The visual images generated by the above recording procedure filled the 19 inch television monitor in the following manner. The head and face were represented, on the average, in the upper 5 1/2 inches of the screen with the torso and neck areas represented in the remaining lower 7 1/2 inches of the vertical dimension of the screen. The upper boundary of the screen met with the top of the speaker's head, while the lower boundary met with the speaker's waist. Such a set of visual-aspect ratios probably made viewing the television image somewhat equivalent to viewing a speaker face-to-face at a distance of approximately nine feet.

Response sheets. Since the dependent measure in Experiment 1 was comprehension, a test had to be designed which would both be a sensitive measuring instrument and be non-intrusive. It was decided that a multiple choice format would be used, since in addition to being easy to administer and score, it would be especially useful for testing specific kinds of semantic information. A model developed by Fillmore (1971) was employed to generate questions because it categorized semantic elements in an utterance as aspects of an event. This was especially well suited to the action descriptions which were to be used as stimuli. Five categories were adopted from the overall set:

1. Agent (the thing that does something in the sentence)
2. Action (what is done by the agent)
3. Location (where the thing is done)
4. Recipient (who the thing is done in relation to)

5. Qualifier (an aspect of one of the above)

Each of the above five categories served as kernels for generating questions. For example, a question might be phrased as: "What the main thing in the statement did was . . . . .?" The possible answers for all questions were carefully constructed so as to measure degrees of comprehension without biasing answers in any direction. To accomplish this objective, all questions used the same answer format. The possible answer types were: (1) the correct answer, (2) an answer which shared many semantic features with the correct answer (e.g., "man" vs. "woman"), (3) an answer which was not an impossible one but which did not share many semantic features with the correct answer (e.g., "drive" vs. "locate"), and (4) three bogus answers which as often as possible sounded similar to one of the first three answer types (e.g., "turn vs. "burn"). The last three answer types were never syntactically or semantically impossible answers.

Thus, for each stimulus item presented to listeners, five individual multiple-choice questions were constructed in the above described manner, each question containing one correct answer and five incorrect answers. Since there were twelve stimulus items, the test consisted of 60 questions which were used to construct an answer booklet (see appendix B).

In addition, each test booklet contained a set of instructions (see appendix A) which served to orient listeners to what they should and should not do. The purpose of the test was explained as a "test of memory." The booklets also

contained two sample items with questions and answers which corresponded to two sample videotaped utterance stimuli. Thus, respondents did have an opportunity to become familiar with the task. Finally, test booklets were arranged so as to contain all five questions for individual stimuli on one page so as to prevent listeners from reading questions prior to the presentation of a stimulus.

Subjects. The respondents were recruited from a basic course in communication at Seton Hall University. Forty persons were tested within Experiment 1 (20 female and 20 male). Subjects' ages ranged from 18 to 24. All were full-time day-session students who were taking a required course. None appeared to have any hearing or vision problems which could be informally detected during the administration of the sample items nor did anyone at any point indicate that he or she was aware of the actual purpose of the study.

Equipment. The visual portion of the presentational stimuli was obtained and presented with the use of a SONY 1/2 inch video recording system, while playbacks were shown on a standard 19 inch television monitor. The aural portion of the stimuli, obtained simultaneously with the above video equipment, was presented with the monitor's volume control kept at a constant level. A white noise background was produced by recording a noise signal from a Grason Stadler white noise generator (901B) onto an audiotape. The noise could then be played back on a standard audiotape playback machine placed next to the television monitor. Figure 2

below shows the arrangements of materials and equipment for both the recording and playback situations, both of which took place in the same room, which contained several tables and chairs, and which measured 6 feet by 9 feet. The room used had one door and no windows and was reasonably free of extraneous noise or visual distractions.

Control of signal to noise ratios. Since the signal to noise ratios were to be manipulated the following procedure was employed. A brief noise-burst was recorded onto the videotape to be used for the experimental presentation just prior to the 12 stimuli. This noise-burst was included for later use in determining the ratios, using portable equipment. Also, the stimulus videotape was used to obtain a printed record of amplitude peaks for the durations of each of the 12 utterances and the brief noise-burst. As can be seen in Figure 3, average amplitudes for the stimulus items were determined by summing the individual values of peaks and taking means, yielding average utterance amplitude values which could then be related to each other and to the value of the brief noise burst that had been recorded on the presentation tape prior to the twelve items' position on that tape.

An RCA sound level meter (WE-130A) was then positioned at a point where subjects would sit in the actual room to be used for collecting data. With the volume level of the television monitor's speaker permanently positioned at one setting (which was a comfortable listening level), it was possible to determine the relative average amplitudes by referring

Figure Caption

Figure 2. Arrangements of furniture and equipment for both recording (a) and playback (b) situations.

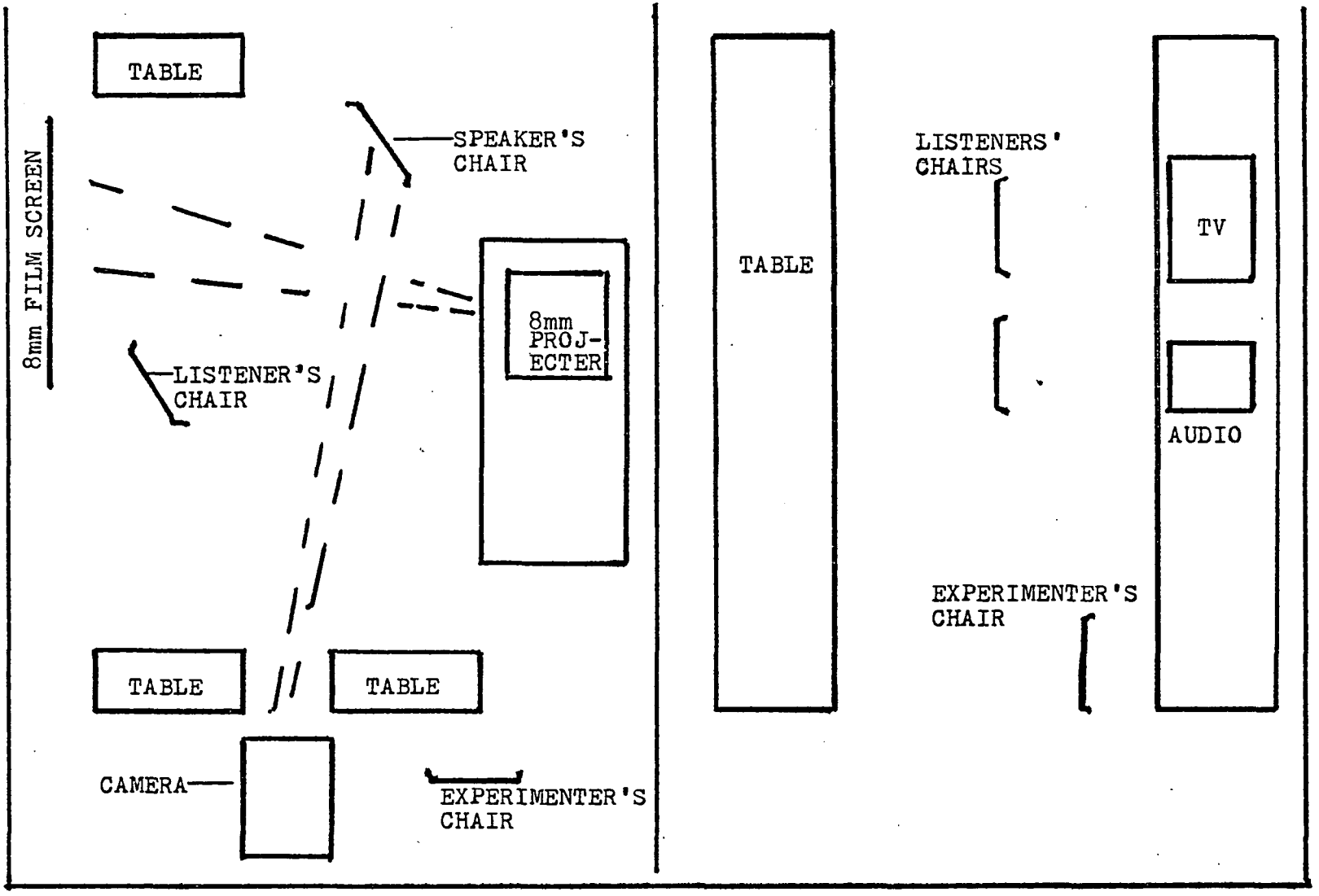
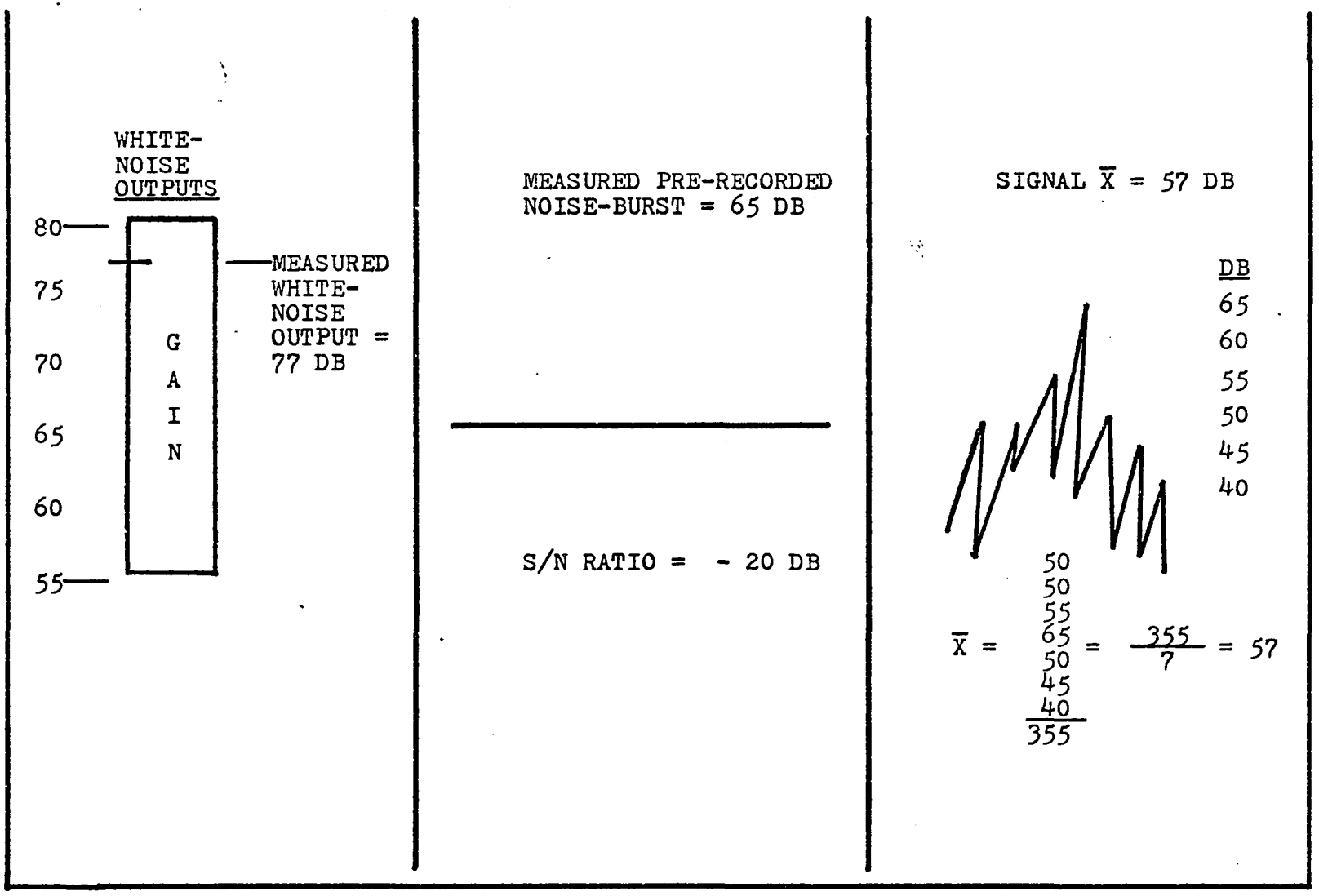


TABLE
SPEAKER'S CHAIR
8mm FILM SCREEN
8mm PROJ-ECTER
LISTENER'S CHAIR
TABLE
TABLE
TABLE
CAMERA
EXPERIMENTER'S CHAIR
LISTENERS' CHAIRS
TV
AUDIO
EXPERIMENTER'S CHAIR

Figure Caption

Figure 3. Procedure for determining the gain-position of noise output to control signal to noise ratios.


WHITE-
NOISE
OUTPUTS
80
75
70
65
60
55
G
A
I
N
MEASURED
WHITE-
NOISE
OUTPUT =
77 DB
MEASURED PRE-RECORDED
NOISE-BURST = 65 DB
S/N RATIO = - 20 DB
SIGNAL $\bar{X}$ = 57 DB
DB
65
60
55
50
45
40
$\bar{X} = \frac{50+50+55+65+50+45+40}{355} = \frac{355}{7} = 57$

to the 12 means calculated from the printed records described above. The amplitude of the white noise background was then determined for all of the points (marked by the manufacturer) on the volume control knob of the audiotape playback machine used for emitting the white noise background, using the RCA sound level meter described above. Thus, by adjusting the gain on the machine playing the white noise to one of its fixed positions, it was possible to control the signal to noise outputs for the twelve items to be used in the study.

Since the above procedure might seem complicated, an example may be helpful. If a -20 db signal to noise ratio was desired for a particular stimulus item whose average amplitude was ten db lower than the db value of the constant noise burst recorded on the presentation tape, the following would be done. The white noise output would be adjusted to a level ten db higher than the level of the noise burst on the presentation tape. The sound level meter would then be used to check the db levels of the recorded noise burst and the output of the noise track, from the fixed distance described earlier. At this point, knowing that the noise output is 10 db higher than the recorded noise burst, and knowing that the recorded noise burst is 10 db higher than the average db value of the stimulus, would determine that the signal to noise ratio of the total presentation stimulus to be -20 db (see Figure 3).

Previous research (Sumby and Pollack, 1954) had indicated a general range for speech intelligibility (-30 db

to 0 db) that would include thresholds for difficulty and ease of comprehension of single words. Using this general range as a guideline, the actual signal to noise values selected were obtained with a pretest using volunteers who independently indicated when they: (1) could not understand the speech but were aware of the vocal presence of the several actual items pretested; and (2) could easily hear the stimulus items. A staircase method, where amplitudes were decreased or increased stepwise, was used to determine the thresholds for the two volunteers' judgements about speech intelligibility. The two subjects' data were quite consistent with with each other, yielding averaged intelligibility threshold values at -8 db and 7 db. Since four ratios were to be used, two midpoints were also selected, leaving four values: -8 db, -3 db, 2 db, and 7 db.

Procedure. In both of the conditions used, subjects would report at their scheduled times to the room where the study took place. Subjects were tested two at a time and were informed that they were participating in a study on memory and that they should try to do as well as they could. Upon entering the room they would be invited into the testing area after some casual conversation with the experimenter about topics unrelated to the study. They would be given test booklets which contained instructions on the front page and would be seated in chairs carefully kept at a constant distance (three feet) from the audiovisual equipment, on all occasions.

Since there were 40 subjects in all, 20 (10 male and 10 female) were randomly assigned to the audio visual condition and 20 (10 female and 10 male) assigned to the audio-alone condition. In the audiovisual condition, all 12 items would be played, in turn, showing the upper half of a speaker's body, which filled up the entire screen area. The speaker's body features were clearly detailed in the picture, while the viewing angle was natural to look at with the speaker looking in the direction of the camera but seated at a slight angle away from the camera (see Figure 2b). Thus, in this first condition viewing the stimuli was not unlike watching a brief excerpt from a broadcast television talk-show. The remaining 20 subjects were assigned to the audio-alone condition, which was identical to the first condition in all respects, except that the video portion was not presented. Subjects in this second condition were told that there had been no picture recorded for their particular task.

In both the audiovisual and audio-alone conditions, the main parameter of the study was regulated by grouping the 12 stimulus items into four groups of three. For example, the first pair of subjects in both of the two main conditions received items 1, 2, and 3 at highest signal to noise ratio (7 db); 4, 5, 6 at the next highest (2 db); 7, 8, 9 at the next highest (-3 db); and finally, 10, 11, and 12 at the lowest (-8 db). A balanced rotation was used in both conditions so that all of the twelve items would be presented equally at all four of the S/N ratios and so that

all subjects would receive an equal number of items in each of the four S/N ratios. In addition, the order of signal to noise levels was rotated so that no serial effects were possible with respect to that parameter.

Each pair of subjects were in both conditions given the test booklet and asked to read the instructions listed on the first page (see appendix B). The instructions asked them to: (1) try as hard as possible to score high on a test of "memory;" (2) not to read any questions before receiving the stimulus statement; (3) answer all questions. The experimenter augmented the instructions by explaining that some items would be easy and some hard to answer. He also explained that every answer counted for one point out of a total score potential of 60.

Once it was observed that subjects understood what they were to do, they were given two sample items to listen to and answer questions about. One sample was presented at the lowest signal to noise ratio and one at the highest. Subjects answered questions similar to the actual questions for the 12 items to be tested and the experimenter checked their answers to enable them to see how well they did on sample items. For all subjects the experimenter would say at this point, "that is in the range of how well others have done on sample items." Immediately after receiving sample items, subjects in both conditions received the actual stimuli at S/N ratios within a particular progression, responding to each stimulus,

in turn, by answering five questions before going on to the next stimulus. The experimenter saw to it that all subjects answered all questions in the booklet. At the end of a testing session, each person was asked to fill in information regarding age, sex, and name on the first page of the booklet. Subjects were debriefed by being told that they would be informed of the specific objectives and outcomes of the study at its conclusion. The experimenter then scored all booklets from the same master list of correct answers.

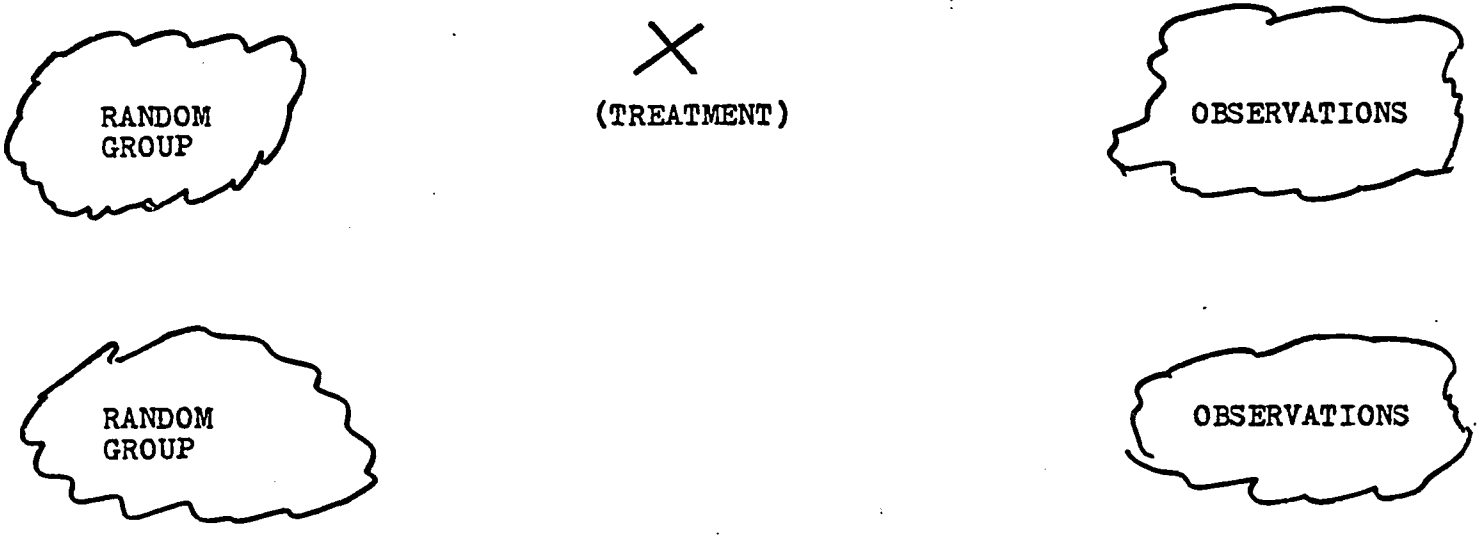
Validity of procedure. In any scientific study the question of general validity of the design is important. That is, can the design assure that the results occurred as they did because of the experimenter's manipulations or did extraneous variables produce the results? Several criteria on the matter are suggested by Campbell and Stanley (1963). Within their schema, the present study used a posttest-only control group design. In Figure 4 below it can be seen that two randomly assigned groups (audiovisual and audio-alone) were observed in exactly the same way with one group receiving the experimental treatment of deprivation of visual cues.

This type of design controls for all types of internal threats to validity which might confound the experimental procedure (such as unequivalent group compositions or subject learning effects). To further secure internal validity, subjects in both conditions were scheduled to report over the better part of daytime hours of weekdays for three weeks to insure that such subject factors as hunger, fatigue, or

Figure Caption

Figure 4. Type design used in present study, according to model provided by Campbell and Stanley (1963).

POSTTEST-ONLY CONTROL-GROUP DESIGN



physiological cycles would be randomly distributed as much as possible. In addition, it was decided that subjects would not participate in both conditions, as their own controls, for two reasons. First, during the early pilot sessions it was observed that when subjects participated in both conditions they quickly came to expect to do better in the audiovisual condition. This expectation gave rise to a self-fulfilling prophesy which was biased in the direction of the main hypotheses of the present study. Second, random assignment of subjects to groups assumes the creation of equivalent groups, in any case.

The concept of external validity concerns the extent to which the tested population is representative of larger populations. Two considerations seemed important: (1) the selection of the subject population; and (2) the effects of the overall method of testing subjects. As for the first consideration, such variables as age, socio-economic group, and intelligence may prove to be interesting in relation to the present topic but did not seem to be crucial limitations for a study of adult dyadic communication abilities of a neo-linguistic nature. On the second, and more important, consideration of test-procedure effects on performance, three things were considered--the attention level maintained by subjects, the use of television as a channel, and the subjects' knowledge that they were being tested.

Some limitation of generalizability to wider communication situations may be warranted due to the use of television in

the data collection because: (1) in general, the television image has a lower resolution (degree of detail) than normal visual images, and (2) in particular, the visual composition employed (aspect ratios) approximated viewing a "live" speaker at about nine feet. Both of these limitations might tend to make the listener's task harder than in normal conversation. On the other hand, viewing television images as well as the subject's knowledge of being tested may have hightened attention. Thus, taken together, the above limitations may, to some extent, cancel each others effects. On the whole, the use of videotape was judged to be an acceptable procedure in consideration of external validity of the design.

RESULTS

The first hypothesis tested in Experiment 1 was that comprehension scores of listeners will be higher when they can see the speaker than when they cannot. At the most general level the results for the 60 item test were that the 20 subjects in the audio-alone condition had an average score of 50% while the 20 subjects in the audiovisual condition had an average score of 61%. This difference was significant at the .01 level of confidence ($t.=2.89$, $df=38$) according to a t-test for independent samples. Thus, the first hypothesis was accepted (see Figure 5a below).

In examining the 60 items of the test more closely, with respect to how much advantage visual cues contributed to test scores, it was found, as was expected, that the large majority of items did benefit, individually, from the addition of kine-

sic cues (44 to 16,) as shown in Figure 5b. This difference was significant ($p < .005$) according to a Sign Test. Since it was established that visual cues did increase comprehension scores, it was decided to investigate just how much of a contribution was possible using the items which had domonstrated themselves to be good differentiators between the two conditions. To evaluate this, the 60 items were rank ordered in terms of how well individual items differentiated the two conditions in terms of the comprehension scores. A subset of 32 items was chosen in this manner and were called "kinesically rich," leaving 28 items which were then called "kinesically poor." The tests that follow will be described in terms of the first 32 items, except in the event that the calculations of a particular test, using the full 60 items, would lead to a different conclusion than would be the case using the 32 kinesically rich items.

Hypothesis 2, stated that the advantage of being able to see the speaker will be greater when the signal to noise ratio is low than when it is high. Figure 6 shows the mean scores for subjects in the two conditions of the study over the four signal to noise values used to collect data.

Differences between the audiovisual and audio-alone scores at the four signal to noise ratios were calculated to show how the visual contribution varied in terms of signal to noise ratios. This relationship is shown in Figure 7.

It can be seen in Figure 7 that the lowest signal to noise value generated a difference of 35% between the audio-

Figure Caption

Figure 5. Results for subjects in both conditions of Experiment 1 for all 60 items. Part (a) shows mean socres and part (b) shows the number of items which benefited from the addition of visual cues compared with the number which did not.

PERCENT CORRECT
0
10
20
30
40
50
60
A
TOTAL NUMBER
0
10
20
30
40
50
B

Figure Caption

Figure 6. Comprehension score averages for both audio-visual and audio-alone conditions. Each point is the average of 20 cases for 8 of the 32 items tested at each of the four S/N values.

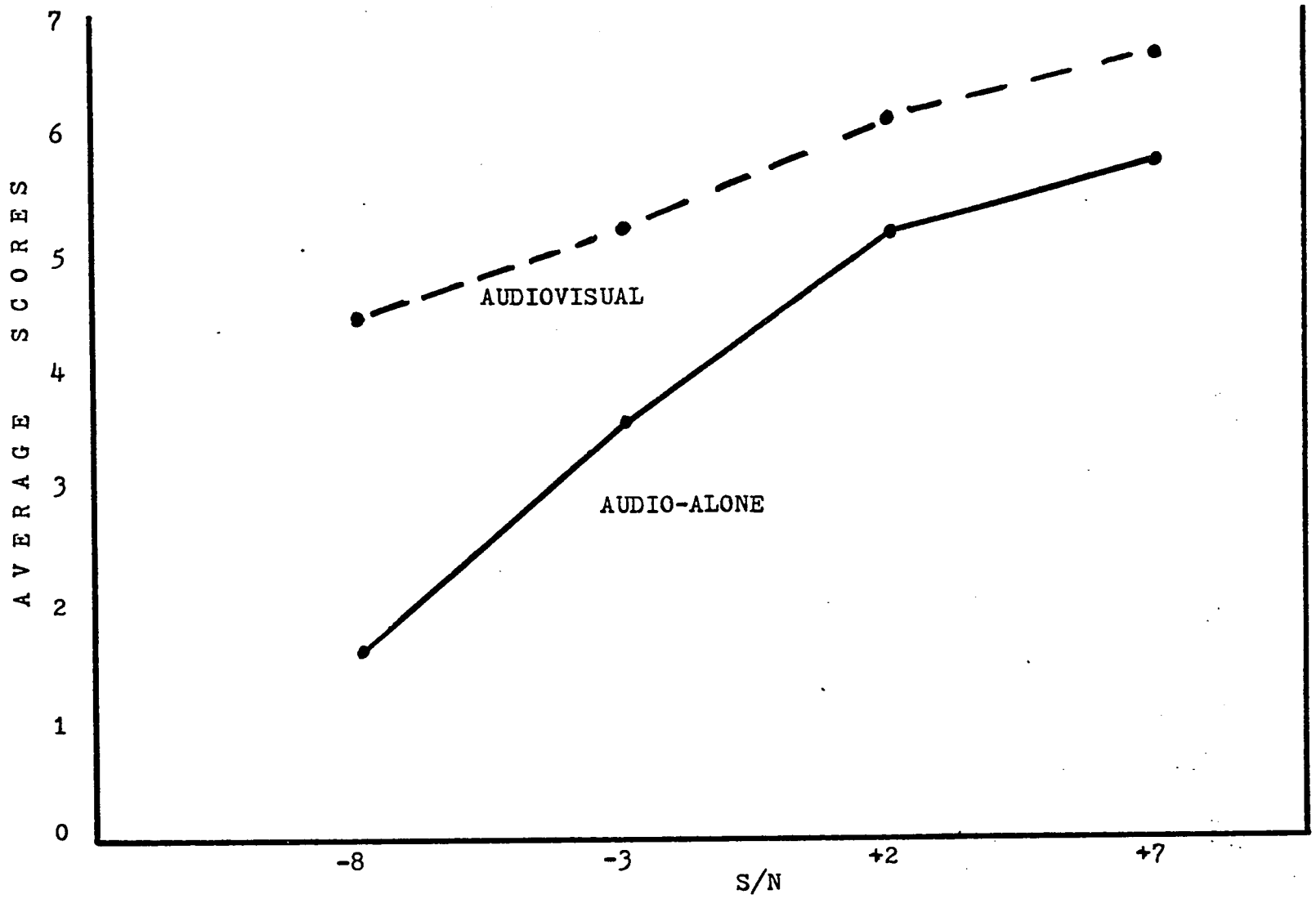
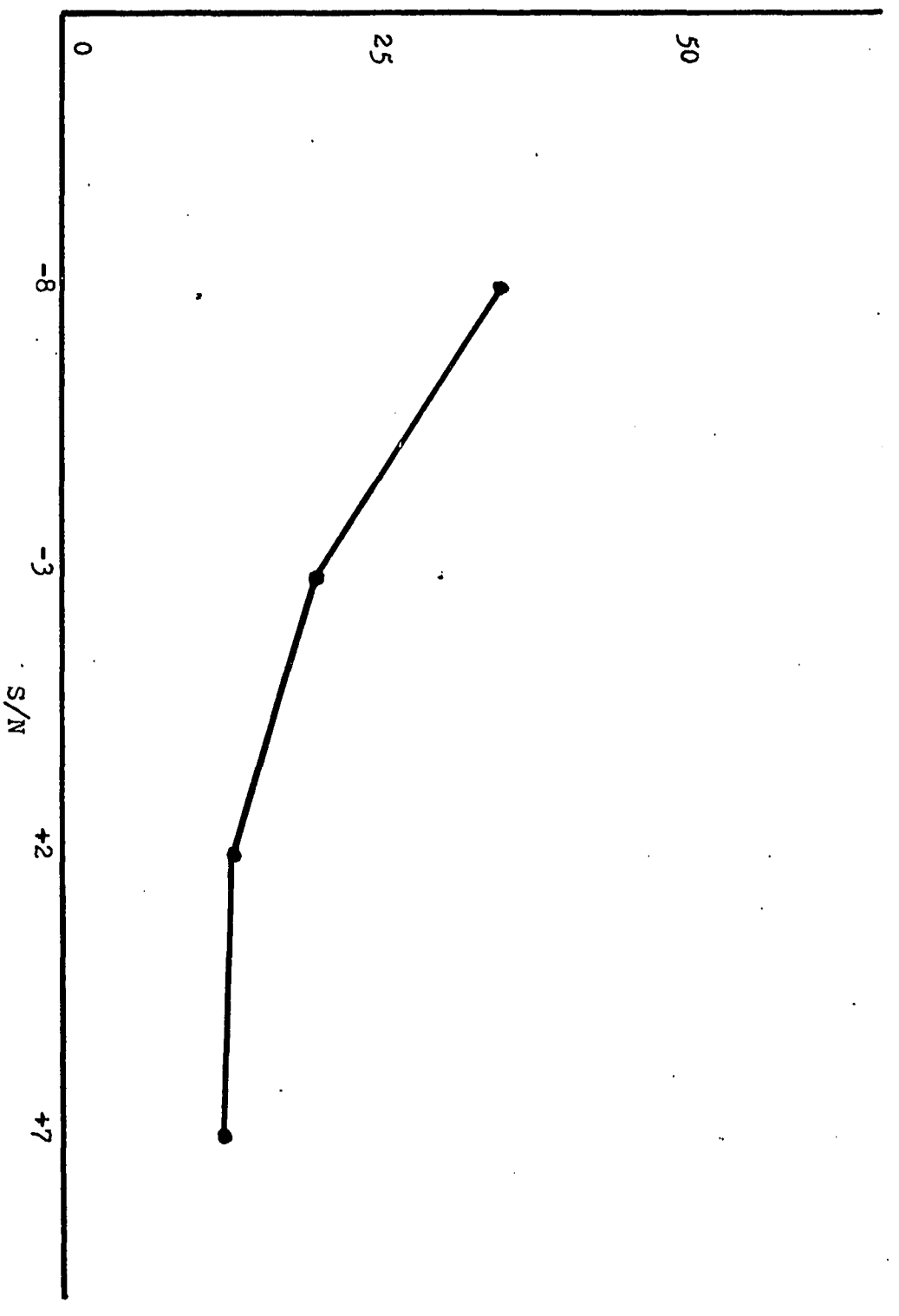

AVERAGE SCORES
7
6
5
4
3
2
1
0
AUDIOVISUAL
AUDIO-ALONE
-8
-3
+2
+7
S/N

Figure Caption

Differences between means in audiovisual and audio-alone conditions over four S/N ratios.


DIFFERENCE IN PERCENT
0
25
50
S/N
-8
-3
+2
+7

visual and audio-alone scores while at the highest signal to noise value a difference of 12% was generated. With the above information, three aspects of the second hypothesis were examined: (1) were means across signal to noise ratios significantly different? (2) were means at particular signal to noise ratios significantly different across the two conditions of the study? and (3) did differences in means across the conditions significantly increase as signal to noise ratios decreased. An analysis of variance for repeated measures was carried out to evaluate the first two of these above aspects and is shown in Table 1 below.

It can be seen in the analysis of variance that the signal to noise parameter was quite effective at influencing comprehension scores and was significant at better than the .001 level of confidence. It can also be seen that the means at the four S/N were significantly different with respect to the addition of visual cues (better than .001). With respect to the third aspect of hypothesis 1, that differences between condition means at the four signal to noise ratios would increase as signal to noise ratio decrease, it can be seen in Table 1 that a significant interaction (better than .048) occurred between the condition and signal to noise variables. A visual inspection of Figure 7 indicates this trend graphically.

So far in the analysis of data, comprehension scores have been evaluated in absolute terms. However, since the room for improvement decreases in the present design as signal

Table 1

Analysis of Variance

Experiment 1

Source of Variance	Sum of Squares	DF	Mean Square	F Test	Significance	Percentage Total Sum of Squares
Condition (AV vs. A)	126.03	1	126.03	58.84	under .001	11.8
Within Subjects	92.37	38	2.43	-----	-----	13.2
S/N Ratio	232.6	3	77.53	37.45	under .001	33.04
Group X Intensity	16.9	3	5.63	2.72	.048	2.4
Intensity X Unit	236.03	114	2.07	-----	-----	33.33
Total	703.90	159	4.43			100.00

to noise ratio increases, it made sense to analyze comprehension scores relative to the room available for improvement when visual cues were added to the aural signal (over the four signal to noise ratios). Figure 8a shows both the absolute differences that had been shown in Figure 7 and the differences between condition means at the four S/N values relative to the available room for improvement. Figure 8b shows in visual terms how relative values were calculated. Here, the X distance reflects a given audio-alone mean, the Y distance the corresponding audiovisual mean, and the Z distance the room for improvement available over the audio-alone mean. R, then, is the relative difference calculated by dividing (Y-X) by Z. Thus, each point in the upper curve shown in Figure 8a was obtained with the above procedure.

It can be seen, then, in Figure 8a that although the visual contribution decreases as signal to noise values increased to make stimuli easier to hear, in relative terms the S/N parameter has a much weaker effect. Consequently, the third aspect of Hypothesis 2 must be qualified; in absolute terms it will be accepted but in relative terms it cannot. The above findings are consistent with the findings of Sumby and Pollack (1954), described earlier.

Hypothesis 3 stated that the types of semantic information tested for in the study would benefit differentially from the advantage of listeners being able to see the speaker. Since the purpose of Hypothesis 3 was to examine the relative benefit visual cues give various types of seman-

Figure Caption

Figure 8. Relative mean differences, absolute mean differences and method used to calculate relative means.

DIFFERENCE IN PERCENT
0
25
50
-8
-3
+2
+7
S/N
PERCENT CORRECT
A
AV
Z
X
Y
R = Z / (Y-X)

tic categories employed in the study, all question items which could be identified as agents, actions, qualifiers, locations or recipients were considered individually. In addition, question items which had been generated from the four original stimulus utterances where speakers spoke on abstract topics were also considered. Because of the nature of the content in these utterances, question items were identified as either: nouns, verbs, complements, or purpose explanations (e.g., "Money is used because...?"). In all, nine categories were examined, using the entire set of original 60 items.

T-Tests were employed to evaluate to what extent seeing the speaker increased respondents' test scores for the items in the above nine categories. Of the nine categories tested, four turned out to be clearly significant: agent, action, noun and verb; one approached being significant: location; and four turned out to be clearly insignificant. However, all nine mean differences across the two conditions (with the signal to noise parameter collapsed) were in the expected direction of a visual advantage. Table 2 summarizes the above t-test analyses.

The third hypothesis was accepted, as four categories clearly differed from four others with respect to the contribution of visual cues. Of interest here, also, is the consistency found between the two different category schemes employed. That is, both agents and actions from the motion questions and nouns and verbs from the abstract questions were

the more successful categories. These categories are obviously parallels of each other. Similarly, recipients and complements were parallels on the negative side.

In summary, the results obtained in Experiment 1 generally support the hypotheses of the study. In addition, an Ex Post Facto hypothesis was generated from two of the original hypotheses. In Hypothesis 2 a kinesically rich set of 32 items were isolated on the basis of having been good discriminators between comprehension scores of the two major conditions of the study. In Hypothesis 3 all 60 items were examined within semantic categories to also assess which items were especially good at discriminating comprehension scores across the two conditions. On a statistical level it seemed quite probable that the categories tested in the last hypothesis were quite able to explain the sorting process that had taken place in the rank ordering done for the testing of the second hypothesis.

Quite expectedly, an inspection of the data showed that of the 32 items which were selected as "kinesically rich" 75% were members of the successful semantic categories of agent, action, noun, and verb. The remaining 25% of the rich set were then members of the less successful semantic categories of location, complement, purpose, qualifier, and recipient. Had such a relationship been predicted, it would have been significant at the .01 level of confidence according to a binomial test.

This last analysis is of more than incidental interest,

Table 2

Summary of T-Tests

Experiment 1

Variable	Mean Percentage-Correct AV	Audio	Differences	T-Values	2-Tail Probabilities
Agent	.69	.55	.14	2.74	.009
Action	.67	.54	.13	3.47	.001
Subject-Noun	.80	.66	.14	2.2	.034
Verb	.72	.43	.29	3.47	.001
Location	.59	.49	.10	1.9	.065
Qualifier	.49	.47	.02	0.36	.723
Complement	.57	.51	.06	1.23	.226
Recipient	.50	.49	.01	.092	.362
Purpose	.68	.65	.03	0.38	.704

since it provides an additional rationale, beyond assessing how functional kinesic cues can be, for having selected the 32 "rich" items for testing the second hypothesis---- specifically, that of demonstrating where and when kinesic cues are especially functional, or possibly non-functional.

DISCUSSION

Experiment 1 sought to measure the contribution of visual cues to the overall comprehension of speech. Specifically, the study sought to measure the increased accuracy (as indicated by a multiple choise test) provided by the listener's opportunity to see the speaker's body movements. In addition, it was expected that noise level and semantic type of utterance would affect the degree to which accuracy increased from the availability of visual cues. The major hypotheses were confirmed and normative measurements noted.

With the above findings in mind, it seems reasonable to say that under normal communicative circumstances, the loss of visual cues makes the task for the listener harder, especially in noisy circumstances. The overall improvement for subjects who were permitted to see speakers ranged from 20% to 40% (and in some cases up to 80% for specific items) over the scores of subjects who were only listening. Since the probabilities calculated were in the proximity of .001 the above findings may be relied on with some degree of confidence. The reasons for the observed improvement can only by speculated about since these factors were not the major focus of Experiment 1.

However, it seems likely that two particular such reasons could very well concern the possibilities that: (1) visual cues operate as a tracking device where the listener is forewarned about when a speaker will begin particular verbalizations; and (2) the co-occurrence of a visual cue with a particular verbal segment may reinforce the perceptual integrity of that segment.

Also, previous research (Dittmann, 1972) indicated that the high information word in an utterance very often receives kinesic stress. This tendency also might serve to focus the listener's attention toward the salient points in the verbal signal, thus insuring that more important aspects will not be missed. Of course an obvious possibility for explaining increased accuracy with visual cues concerns the semantic redundancy provided by these cues with respect to verbal segments. Finally, one indirect explanation concerns the concordance of visual cues observed previously (Kendon, 1972) with prosodic segments. This concordance may serve to increase the stimulus strength of prosodic elements which in turn facilitate the perception of verbal segments.

Of major importance in Experiment 1 was the manipulation of signal to noise ratio. As expected, the noisier the circumstances the more useful were the visual cues, in absolute terms. In fact, subjects frequently expressed frustration to the experimentor when attempting to respond to the noisiest level (-7 db). However, it was precisely at this level that they most utilized visual cues to record quite respect-

able comprehension scores (50% in the audiovisual condition) considering the obstacles subjects faced at this level.

However, when accuracy scores were considered relative to the available room for improvement an interesting trend was found. That is, the easier listening conditions enabled subjects to perform equally as well as did the more difficult listening conditions. This was opposite to what had occurred in terms of absolute score means. One explanation that seemed plausible is that the ceiling operating at the easier signal to noise ratio listening conditions masked two opposing forces on the data. That is, on one hand easier to hear circumstances suffer the decreasing availability of room for improvement factor, when absolute score means are considered. However, on the other hand, the more difficult to hear circumstances suffer a loss in the average cue value within the message. This is so because when information about a message is generally low, some cues which cannot be related to anything already known about the identity of the message do not attain a useable information value. Thus, the mutal cancellation of both forces may only be observable when considering relative scores.

Also, of major importance were the differential performances of the nine semantic categories tested in Experiment 1. It had been expected that some categories would benefit more from visual cue availability than others, since previous

research had indicated that some types of linguistic information are more kinesically codable than others (Graham and Argyle, 1975). The finding that agents and actions were supported strongly by kinesic cues seemed not surprising on an intuitive level, since they seemed well suited to the action "language" potential of body movements. However, the finding that the subject-noun and verb questions used to te[illegible] abstract items benefited as much as the first two with added visual cues was somewhat of a surprise.

Two possible reasons may explain the above findings. First, it was earlier suggested that one manner in which visual cues may support speech comprehension is by visually stressing the more important parts of the verbal utterance. With this explanation, semantic content as a variable is not important. A second possible explanation is that abstract verbal material can be kinesically supported semantically through a type of metaphorical coding where, for example, some abstract sentence-subject is given visual spatial dimensions with hand gesticulation (e.g., if the subject of a sentence were the phrase "political power" the hands might assume claw-like dimensions and move as if uplifting a vast quantity of earth).

Thus, although the data do support the notion that different semantic categories benefit more from visual cues than do others, this tendency is more complicated in nature than at first thought. What was demonstrated in Experiment I was that very major types of semantic categories (i.e., agent,

subject, action, verb) benefited more from visual support than the relatively secondary semantic categories (i.e., qualifier, complement, recipient, purpose, and location). As said earlier, whether the former categories were more codable or simply received greater visual stress will have to await further study. One rather risky premise based on the limited set of categories tested was that the poorer categories were more transformationally complex (e.g., qualifier). Of course the present design was not constructed in any way to address this premise.

Two additional variables turned out to be unimportant to the present study--sex of subject and length of utterance. The first did not show a significant difference for either speakers or listeners. However, earlier studies have shown that females can at times decode nonverbal signals regarding emotions more accurately than can males (cf. Zuckerman et al, 1975). Presently, the possibility that this tendency would generalize to descriptions of spatial motions was not found. However, future studies might investigate possible sex differences with a greater variety of semantic topics to be encoded and decoded. Despite the absence of significant findings in the present data, the topic seems an interesting one because of the social and biological sex role differentiation with respect to the rearing of children--especially children's acquisition of a communicative competence.

The second variable concerned the effects of the length of stimulus-utterances on comprehension scores. It had been

expected that the addition of visual cues would be related positively or negatively to overall length of utterance. However no trend in either direction was found. Possibly other variables of a syntactic and/or semantic nature masked such a trend; consequently, little can be said of the length variable within the data collected for the first experiment.

Probably the most important limitation of Experiment 1 concerns the role of lipreading in the visual component of communication. That is, the present study made no attempt to distinguish between what portion of the visual contribution was due to lip cues and what portion was due to body movements other than in the lip area (the two are conceptually different media of communication).[7] It seems reasonable to say that the important question is not whether the improvement was due to one or the other, but rather the important question concerns measuring how much improvement is generated by each alone and by both together. While the next segment of the present text addresses this question, the data collected for this segment can probably be used to demonstrate that lip cues by themselves could not have accounted for the total improvement due to visual cues. The reason for making this assertion concerns the differential improvement found with respect to the nine semantic categories tested. Lip cues probably assisted equally in all categories (assuming an equal distribution of visible and nonvisible articulatory movements within verbal segments on the order of 25 words each); however, the nine semantic categories performed unequally with

respect to the improvement rendered by visual cues.

Thus, it was deemed important to evaluate the role of lip movements in relation to the overall contribution of visual cues to speech comprehension. This evaluation was made in Experiment 2, which follows.

EXPERIMENT 2

A major purpose of Experiment 2 was to establish that the visual contribution to listener comprehension found to operate in Experiment 1 could not be solely attributed to lip cues. Two considerations seemed to warrant examining this contention in detail.

First, the indirect assessment made at the end of Experiment 1 (where it was shown that different types of semantic content benefited differentially from the addition of visual cues) suggested that the overall visual contribution to comprehension varied in extent of its utility much more than could be accounted for by momentary variation in lip cue utilization made by listeners.

Second, a study by Woodward and Barber (1960) found that naive normal-hearing listeners could only discriminate lip cues in terms of four general classes of cue types when they were shown silent-filmed recordings of two-syllable nonsense words. They could not identify the individual 22 consonants' lip cues that were tested for on an individual basis. Also, the Sumby and Pollack (1954) study described earlier suggested that listeners can only utilize lip cues to identify words when the to-be-perceived word is narrowed down to

a set of from 8-256 possible choices. In normal conversation, then, it would seem that lip cues are probably facilitative to comprehension only from time-to-time, depending greatly on specific speech contexts.

Certainly, if lip cues are not the sole visual facilitative factor, kinesic cues, by inference, must also participate as a factor in the listener's improvement in comprehension due to display of visual cues. A second purpose, then, of Experiment 2 will be to make some assessment of the contribution kinesic cues made for listeners viewing the 12 stimuli, apart from the utility of lip cues.

Two reciprocal hypotheses were formulated to test the above contentions:

1. Lip cues alone cannot account for the increase in comprehension found when listeners are permitted to see the speaker.

2. Conversely, the kinesic cues that remain after parts of the visual component of speech are electronically removed can be used to demonstrate increased utility for listeners' speech comprehension.

Method [8]

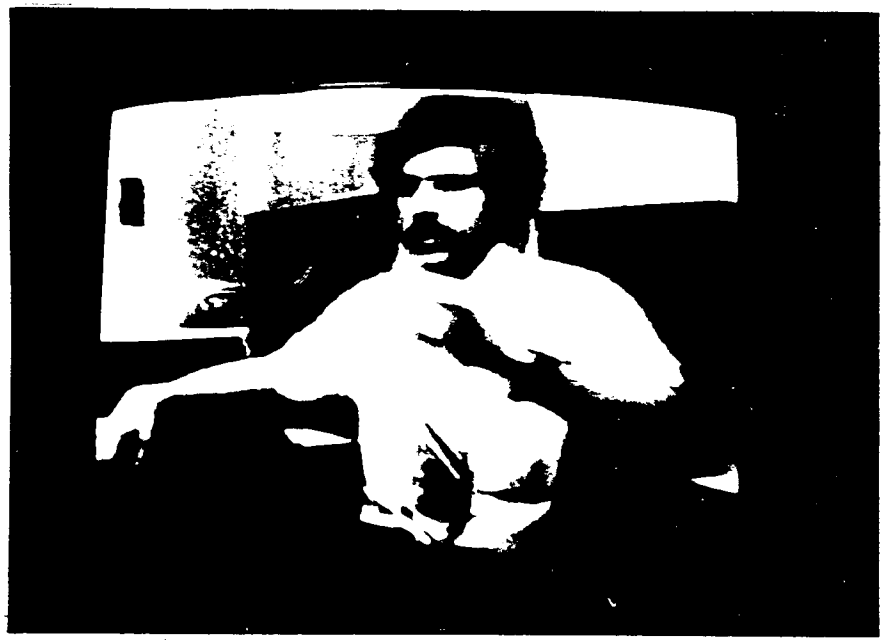
Audiovisual stimuli. The presentation stimuli for Experiment 2 were identically the same set of 12 speaker utterances which had been produced for Experiment 1 by having speakers describe brief filmed actions to a second person. However, in order to control for lip cues, the original stimuli were visually altered by adjusting the brightness and contrast controls on the television monitor used for presenting

stimulus items. Settings on these two controls were found that effectively blocked out the general mouth and eye cavity areas of the face while leaving the grosser features of the speaker's image visible (head, forehead, cheeks, shoulders, arms and hands). Figure 9 illustrates the visual image presented in the contrast-reduced condition compared with the normal visual image used earlier. This electronic manipulation had been planned for with the lighting arrangement that had been used to record the original stimulus set. The lighting had been controlled so that with normal brightness-contrast adjustment all body details were relatively clearly depicted while with lowered brightness-contrast adjustment any body detail which was relatively low in visual intensity (but visible) would become washed out in shadow (the difference in video intensities between the normal and lowered adjustment were about 1.0 volts and 0.4 volts, respectively).

Other methods that were available to the experimenter were considered with respect to controlling for lip movement. For example, speakers might have been required to wear masks which would block lip cues from view or might have been trained to restrain their lip movement. However, it was reasoned that either of these techniques would intrude on speakers' kinesic and verbal performances to such an extent that these performances would be significantly different from the performances used in the free-movement audiovisual condition which was studied earlier and from which data would be compared with data collected presently.

Figure Caption

Figure 9. Comparative illustrations of visual images presented in audiovisual condition and audiovisual (altered) condition.



Audiovisual

Audiovisual
(altered)

In order to be assured that the above lip control that was adopted would be perceived by subjects in the way intended by the experimenter, two volunteers were asked to view each of the 12 stimuli at five different brightness-contrast adjustments which ranged in small steps from a point where lip movements could easily be seen to a point where they clearly could not be seen. Both volunteers independently rated items at the five adjustment levels, responding in terms of how well they could see the lip area (very well, ok, slightly, just barely, and not at all). The brightness-contrast adjustment selected reflected a midpoint between their combined judgements of "not at all" and "just barely." Both volunteers independently agreed that this setting did not show the lip area in any appreciable detail. In addition, the experimenter was satisfied that the adjustment did in fact mask lip movements.

Response sheets. The same response sheets constructed for Experiment 1 were used, where each of the 12 items were tested for listener comprehension with five multiple choice questions, yielding 60 in all. In addition, instructions to subjects, sample items, and requested subject information were all included exactly as had been the case earlier.

Subjects. Twenty volunteers from a basic course in communication at Seton Hall University served as subjects, none of whom had participated in, or knew anything about, the first experiment. Subjects' ages ranged from 18 to 28. All were day-session students who were taking a required

course. None appeared to have any hearing or visual problems which could be informally detected during the administration of the sample items nor did anyone indicate that they were aware of the specific purpose of the study.

Equipment. The equipment used consisted of the very same SONY 1/2 inch video recorder, audio playback machine, Grason Stadler noise generator, and RCA sound level meter that had been employed in Experiment 1.

Control of signal to noise ratios. Four signal to noise ratios were used in Experiment 2 (-8db, -3 db, 2 db, 7 db) which corresponded to the four levels used earlier. Again, these ratios were maintained: (1) by fixing the volume control of the television monitor at one setting; (2) by referring to the printed record of stimulus amplitudes (and brief noise burst); and (3) by using the sound level meter to determine the amplitude of the white noise source relative to the amplitude of the brief noise burst recorded on the stimulus tape just prior to the set of 12 items. Thus, by adjusting the gain on the machine playing back the white noise source to one of its marked positions (whose amplitudes had been measured with the sound meter) it was possible to present stimuli at the desired signal to noise ratio.

Procedure. The group of 20 subjects were randomly assigned to one of four signal to noise ratio rotations (with the 12 items grouped into four groups of three) to assure that all 12 items would be presented equally at the four ratios and to control for order effects, as had been done in

Experiment 1. Subjects were tested in groups of two and three at a time. They would arrive at the testing area and be invited inside and then be asked to be seated. Booklets were then given out and each person asked to read the instructions. As done earlier, the experimenter would orient subjects regarding answering questions and then administer sample items. Once it had been observed that subjects understood what they were to do, testing began. Here, the same procedure was followed at all times: the first stimulus would be presented, the equipment stopped, and the subjects given time to answer all questions. All 12 items would be tested, in turn, in the same manner.

Lastly, the issues about experimental validity of design were the same for Experiment 2 as they were discussed in terms of Experiment 1.

RESULTS

The general question addressed in Experiment 2 concerned whether lip cues alone can account for the overall improvement in listener comprehension derived from the visual component of speech--or whether kinesic cues must be included also as a significant factor in the total improvement. If subjects viewing the contrast-attenuated stimuli could score higher on the comprehension test than subjects who could only hear the 12 speech stimuli, then lip cues alone did not account for the total improvement, since they were not present. In addition, such a finding would strongly suggest (although not prove) that lip cues can not alone account for compre-

hension improvement for listeners in normal face-to-face communication situations.

The average score of subjects viewing the 12 stimuli in the brightness-contrast reduced condition was 60% correct for the same set of 32 question items that had been selected for study in Experiment 1. This mean compared to an average score of 71% correct in the normal audiovisual condition and 48% correct in the audio-alone condition. Figure 10 shows means at the four signal to noise ratios within all three presentation conditions. As can be seen, all three functions vary in the expected directions with respect to the increase in noise and the increase in amount of visual cues presented. That is, as greater amounts of visual cues are added, comprehension scores increase and as signal to noise ratio increases comprehension scores also increase.

An analysis of variance for repeated measures was performed to test differences in means over the three conditions and four signal to noise values. As can be seen in Table 3, mean differences in the different conditions were quite significant as were mean differences across the signal to noise parameter. In addition, individual t-tests were computed among the various combinations of means. The results of these individual tests are shown in Table 4. It should be pointed out that since these individual tests all concerned expected directional outcomes, they could be regarded as single tests with respect to the selection of confidence levels.

Because the analysis of variance had shown a significant interaction effect between condition (audiovisual, audiovisual-altered, and audio-alone) and signal intensity, separate analyses of variance for repeated measures were performed for data in the two lowest signal to noise ratios and in the two highest signal to noise ratios. The results of this second analysis were found (as could be expected) to be virtually identical to what is shown in Table 3, for the entire data set, with the one exception being that no significant interaction effects between presentation-condition and signal intensity remained when considering either half of the function shown in Figure 10. Table 3 shows the non-significant interaction effect when testing data in either half of the function.

Because the two halves of the overall function were found to have non-significant interaction effects, normalized representations of the function were constructed graphically. These two normalized halves of the overall function are shown in Figure 11. The impression that emerges from this diagram is that the noise parameter significantly effects the extent to which the reduced-contrast visual cues were utilized by listeners. It can be seen that listeners most used these visual cues when the acoustic signal was poorest.

It would seem reasonable to say that the above analyses of data suggest that lip cues did <u>not</u> in fact account for all of the improvement in listener comprehension due to the display of visual cues accompanying speech. And so, the two main hypotheses tested in Experiment 2 were tentatively

Figure Caption

Figure 10. Mean scores for audiovisual (altered) condition over the four signal to noise values. Also shown are mean scores for normal audiovisual and audio-alone conditions that had been found in Experiment 1.

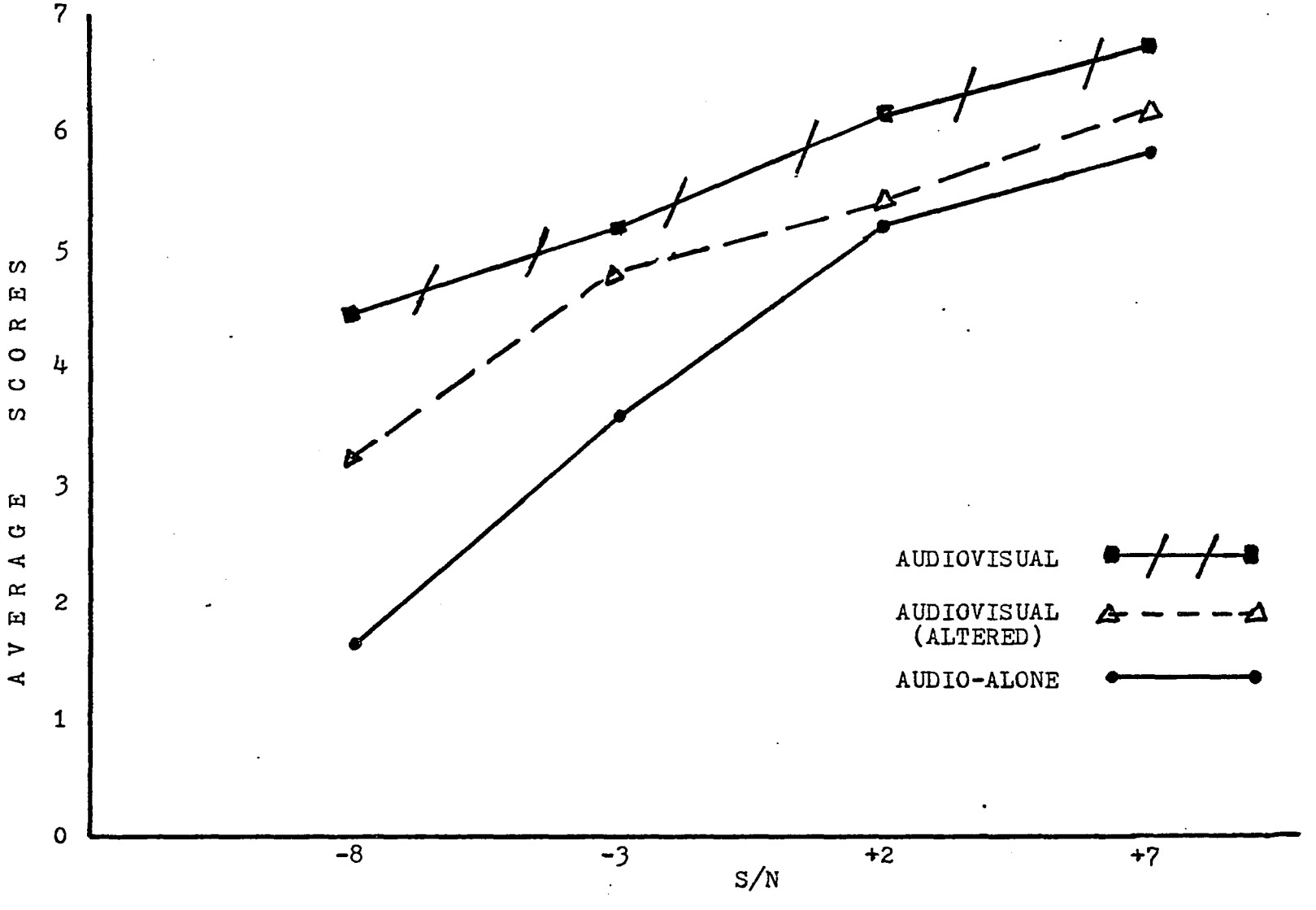

AVERAGE SCORES
7
6
5
4
3
2
1
0
-8
-3
+2
+7
S/N
AUDIOVISUAL
AUDIOVISUAL (ALTERED)
AUDIO-ALONE

Table 3

Analysis of Variance

Experiment 2

Source of Variance	Sum of Squares	DF	Mean Square	F Test	Significance	Percentage Total Sum of Squares
Condition	97.91	2	48.95	23.29	under .001	11.8
Within Subjects	119.84	57	2.10	-----	-----	14.44
S/N Ratio	339.57	3	113.19	78.38	under .001	40.92
Group X Intensity	25.26	6	4.21	2.91	under .01	3.04
Intensity X Unit	247.17	171	1.45	-----	-----	29.79
Total	829.74	239.	3.47			100.00
Group X Intensity	S/N (+7, +2)	2		2.11	n.s.	
Group X Intensity	(-3, -8)	2		2.62	n.s.	

Table 4

Individual T-Tests over Four

S/N Ratios and Three Conditions

Condition				Value of T-Test (df = 38)		
Signal to Noise Value	(1) Audiovisual (normal)	(2) Audiovisual (altered)	(3) Audio-Alone	(2-Tail) 1-2	2-3	1-3
-8 db	$\bar{X}$ = 4.4 SD = 1.47	$\bar{X}$ = 3.5 SD = 1.21	$\bar{X}$ = 1.65 SD = 1.6	t=2.71 p < .01	t=3.57 p < .001	t=5.67 p < .001
-3 db	$\bar{X}$ = 5.2 SD = 1.3	$\bar{X}$ = 4.8 SD = 1.1	$\bar{X}$ = 3.6 SD = 1.5	t=1.06 p < n.s.	t=2.84 p < .008	t=3.58 p < .001
2 db	$\bar{X}$ = 6.15 SD = 1.0	$\bar{X}$ = 5.4 SD = 1.3	$\bar{X}$ = 5.2 SD = 1.2	t=2.04 p < .048	t=0.5 p < n.s.	t=2.63 p < .01
7 db	$\bar{X}$ = 6.75 SD = 1.0	$\bar{X}$ = 6.2 SD = 0.8	$\bar{X}$ = 5.8 SD = 1.4	t=1.87 p < .07	t=1.1 p < n.s.	t=2.45 p < .02

accepted. A preliminary estimate of the two visual modes' individual contribution to the total effect would appear to be about 50% and 50%, respectively. However, a number of factors need to be considered before attempting a reasonable estimate of the role played by the two modes in more natural circumstances. These considerations are taken up in the discussion, which follows.

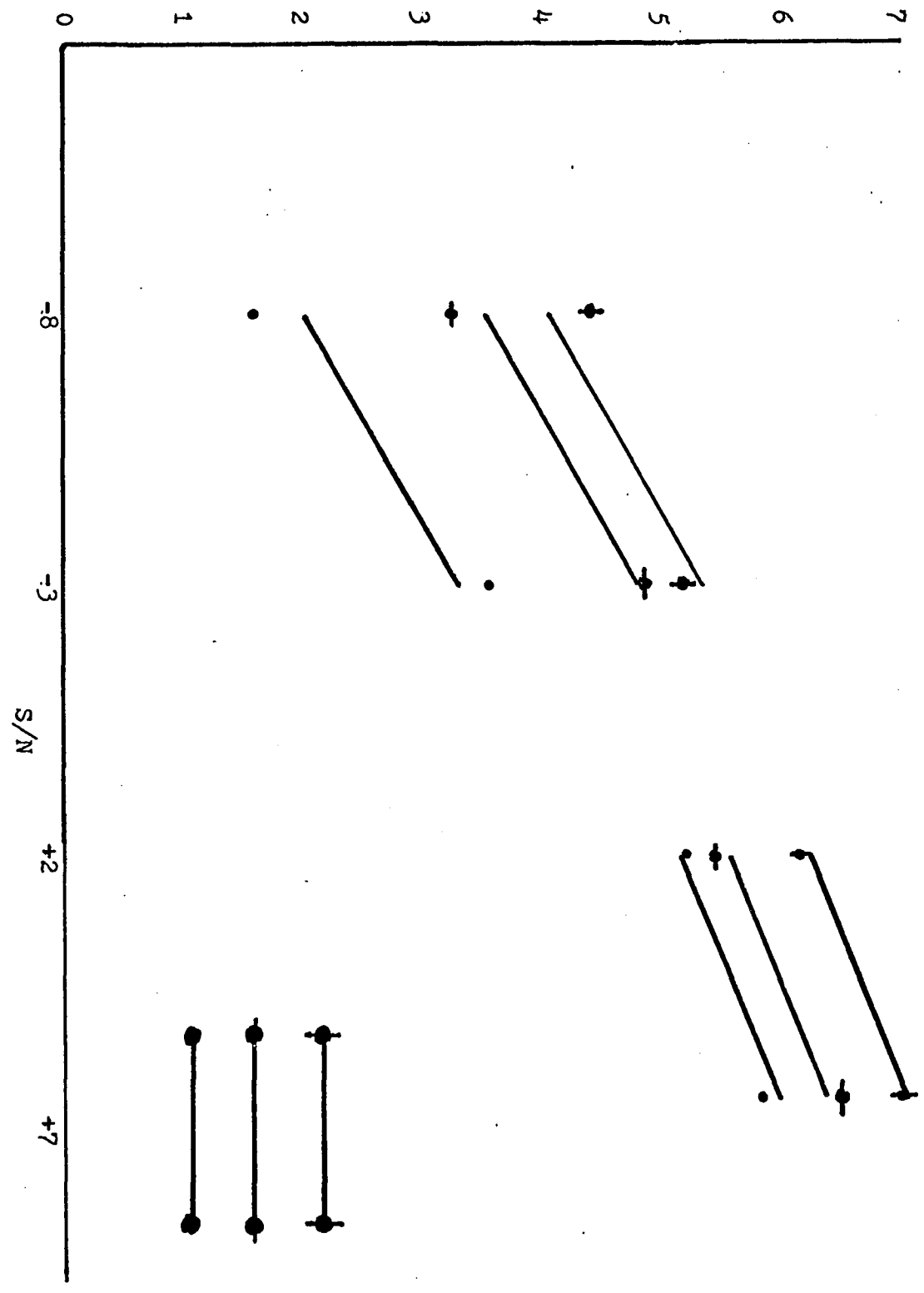
DISCUSSION

Before analyzing the significance of the above findings, one point needs to be made about what the listener likely attends to when processing audiovisual speech stimuli. Individual listeners may attend to the visual component as an overall gestalt or may attend to specific visual cues more closely than to other visual cues. Assessing which cues a listener was attending to when responding to test items in the present study was not an objective of the design used and would likely comprise a major research undertaking in its own right. Rather, the present discussion will focus on the implications of the data analyzed above, with respect to potential utility of specific visual cues to speech comprehension, given that such cues are in fact attended to. With this reservation in mind, several things can be said.

Apparently kinesic cues can make a sizable contribution, as suspected, to the overall utility of visual behavior in the communication of verbal concepts, as did the partial array of extra-labial kinesic cues in the contrast-reduced stimuli tested presently. This partial set of kinesic cues proved

Figure Caption

Figure 11. Estimate of the general function in both relatively quiet and relatively noisy circumstances. Points indicate actual means and lines indicate estimate.


AVERAGE SCORES
0
1
2
3
4
5
6
7
-8
-3
+2
+7
S/N

to account for roughly 50% of the total improvement which had been found to occur when testing the complete set of visual cues compared with the audio-only communicative situation. And it was especially at the lower S/N ratios that the partial set of kinesic cues were especially useful (60-65% of the total visual improvement). Such a trend may indicate that listeners pay different amounts of attention to the various types of communication cues available, depending, in part, on the overall ease or difficulty of processing such cues and depending, in part, on how rich the cue pool is at a given time.

Of course, the major problem in specifically stating what the individual share of the two visual modes (within the total visual contribution) was in the present experiment is that when part of the normal visual stimulus is artificially removed (e.g., mouth and eye cues) the result is that the remaining parts of the visual component may receive listener attention-patterns that vary from the attention-patterns those above remaining parts of the visual component receive under normal viewing conditions. This problem would exist even if only lip cues were experimentally removed from the visual component.

In making some assessment, however, it seems reasonable to assume that the human listener is adaptable in making use of cues which bear communicative information. This assumption is supported by the tendency for listeners to make increasing use of the partial set of kinesic cues (and for that

matter, the entire set in Experiment 1) as more and more acoustic noise was added to the speech signal. That listeners do this same sort of adapting to available cue conditions in normal communicative conditions is further supported by virtue of the fact that the reduced-contrast condition did, in a sense, recreate a natural communicative context involving shadow obscuring visual details. This context is a common circumstance in day-to-day communication, as for example when conversing in a darkened room or on a poorly lit street at night.

It may also be reasonable to assume that the diagram (in Figure 11) showing the individual contributions of the two visual modes to listener comprehension over-represents the individual role of lip cues in the total process, because of the kinesic cues (mouth and eye areas) which were also lost along with lip cues. These additional kinesic cues may well represent some of the improvement due to restoring the complete visual component. If this is the case, a more accurate representation of the share division might approach a 2:1 ratio favoring kinesic cues--given speaker and speech-content samples representative of the samples studies presently. But beyond estimates in terms of numbers, the important implication of Experiment 2 is that extralabial visual cues are at least as important (and possibly more important) than are labial-visual cues for listener comprehension level).

One final point concerns an explanation of why kinesic cues were more valuable to the listener than were lip cues

when noise levels were increased. It may be that listeners come to rely on kinesic cues more than on lip cues in difficult-to-hear situations because of the different natures of both types of visual cues (assuming that "fluent" kinesic cues are present). That is, kinesic cues are more often iconic in nature and might require less intense learning for their occassional utilization in noisy circumstances than do lip cues. That is, lip cues are free to indicate only that a given speech sound has been encoded. However, a kinesic cue may indicate that a given idea has been signaled, or a given nuance intended, or a given syllable emitted. Lip cues are, on the other hand, very much constrained by the prerequisites of articulatory movement, functioning solely at the phonological level while kinesic cues may function at several levels (e.g., semantic, paralinguistic).

At low noise levels, however, the contrast reduced stimuli may have been under-used because the relatively clearer-to-hear speech stimuli were relied on more by listeners, who were at these noise levels already comprehending well. It may be that listeners payed less attention in this situation to the especially low-resolution visual image presented to them here.

CONCLUDING REMARKS REGARDING EXPERIMENT 2

In future attempts to dissect the modal components of the overall assistance rendered by body movements to speech comprehension, a more sophisticated technical procedure will have to be employed. Very likely a luminous

material will have to be used to control which areas of the body are actually recorded in detail. (i.e., lip areas or all other areas). With this type of procedure, the precise weight of both visual modes can be measured both separately and together.

That perceivers can process such visually "abstracted" body movement cues has been documented (cf. Johansson, 1975). The author film recorded people with small battery-operated lights placed on key movement areas of the body (e.g., arms, feet shoulders, etc.). The entire scene was filmed in darkness so that only the light bulbs were observable. When actors remained still, all that would be perceived by viewers were clusters of lights. However, when actors began to move, not only were their forms recognizable as human, but their actions and even their emotions became more and more identifiable as more and more lights were added, enabling images of body movements to become increasingly synchronized.

GENERAL DISCUSSION

Although the findings of both Experiment 1 and 2 were briefly discussed, respectively, in earlier sections, a more detailed analysis will be made now for the purpose of appraising the significance of the two studies major outcomes to future research in the area and to possible applications of the present (and future) findings in relation to both normal and problematic functioning of face to face communication. Specifically, what follows here will be: (1) a consideration of what variables have and have not been accounted for in de-

termining how functional the role of kinesic behavior is to the processing of speech; (2) a review of some of the basic theoretical issues mentioned in the introduction section in light of the present findings; and (3) the postulation of a conceptual framework (in the form of a model) of the present research domain to enable the later making of several suggestions regarding possible future study and applications.

Apparently, the potential contribution of kinesic cues to speech comprehension is quite variable and consequently, describing the potentially many circumstances which determine whether its utility is high or low is an interesting problem. In the present study, visual cues were found to enhance comprehension as much as 35% for specific types of content (more than half of this improvement due to visible body movements other than in the lip area). One idea partially developed earlier was that utility would be high when specific speech content was especially well suited to kinesic coding. Although the meaning of a particular kinesic act and an accompanying speech segment may at times be equal and isomorphic, this probably occurs only occasionally. What seems more likely is that kinesic cues are more often concordant with specific aspects of speech segments such as with the specific semantic or syntactic features of a particular segment (e.g., if one were to move one hand several inches from left to right in front of the body while saying, "the car crashed into the haystack," the hand movement could be said to mark one particular semantic feature of the verb "crashed into,"

specifically lateral movement.

So far, the variables of semantic content and noise have been related to the variance in comprehension performance entailed by adding visual cues to the speech signal. In addition, sex and length of utterance were also mentioned earlier. However, a number of other variables are probably important, although they were not included in the designs of the two studies undertaken here. Basically, these additional variables can be associated with one of three conceptual areas:

sender receiver variables,
message variables,
situational variables,

Sender-Receiver variables. The most obvious variable here concerns the communication skills of both the listener and speaker. For example, speakers very likely vary in what might be called "kinesic fluency." Since the eight speakers varied in their ability to benefit from the kinesic cue availability to listeners, it may have been that the more successful ones were more kinesically fluent. One previous study attempted to explore this matter (Baxter and Winters, 1968), with partial success, by employing a personality measure. The authors suggested that kinesically fluent speakers may be those individuals who maintain a relatively diverse set of distinctions (or cognitive map) about their perceptions of the world. It is possible within this line of thought to speculate that people who use more conceptual cate-

gories in thinking and communicating about their more complex view of reality would find gestural markers useful in helping to keep their spoken thoughts in order. In addition, a number of other speaker variables might play a role in accounting for variability in kinesic cue utilization; these would include such things as age, educational level, intelligence, and social strata.

Two other related speaker variables seem especially deserving of mention here--motivational level and level of arousal. Since bodily activity is closely related to these two variables, it seems reasonable to suppose that when a person is highly motivated (or in a highly aroused physiological state) in some communicative situation, the resulting increased bodily activity will automatically be incorporated within any messages being emitted. Such a relationship can be theorized on the basis of Condon's work (1966, 1974) on body-motion/speech synchrony. Such synchrony is certainly below the level of conscious control (under normal circumstances) and may have a neurological explanation.

Two crucial listener variables concern degree of attention and focus of attention, since when attention falls off, comprehension must fall off. However, with respect to the utility of visual cues, several outcomes may result depending on whether attention falls off either the vocal or visual cues, or both. If, for example, attention momentarilly fell off a particular verbal segment, but not from the accompanying visual segment, the resulting level of comprehension might

not change. None-the-less, listener attention is very likely a rich source of variance which very likely interacts with the listener's attitude toward the content and with such things as listener fatigue. Also, the listener focus of attention (i.e., what cues are attended to) may be important in further studying individual differences in kinesic utility.

Finally, two variables concerning the speaker and listener together seem important enough to mention briefly. The first concerns how familiar they are with each other, since comprehension accuracy may increase with greater interpersonal familiarity. This variable may be especially important with data collection in such designs as the one used presently where listeners are shown one very brief taped excerpt of a speaker whom they have never listened to before. The second speaker-listener variable concerns interpersonal attraction. Again, comprehension accuracy may be affected by this index of dyadic involvement.

Message variables. Probably the most important message variable affecting the utility of kinesic cues concerns the codability of utterance content. That is, intended content has a linguistic codability potential and a kinesic codability potential. Previous research (Graham and Argyle, 1975) has indicated that kinesic utility increases for items with a low verbal codability potential with respect to the description of two dimensional figures. Certainly the kinesic codability potential is equally important with respect

to utility, as, for example, in the above case where it may be presumed that geometric figures have a high kinesic codability potential. A second message variable concerns what is known by listeners about the present message from previous messages exchanged in a given conversation or from knowledge of the topic of conversation. Since these variables undoubtedly influence the accuracy of comprehension, in general, they may possibly influence the kinesic utility factor in particular.

Situational variables. The situation in which a conversation takes place obviously will affect the content expressed. But it may also affect the style of communication, especially the differential use of various channels of expression. For example, the difference between giving directions to one's house to a stranger as opposed to mentioning how one goes home to a friend who knows the route may well affect how much stress is placed on the visual channel for equivalent verbal content. Similarly, one may be lecturing on a topic to a learning group as opposed to an uninterested colleague, using two very different nonverbal (and verbal) styles. Both of the above examples illustrate how the situation the speaker and listener are in may affect the verbal-nonverbal output relationship, with the added possibility that kinesic utility will be dramatically affected with respect to speech comprehension. Additional examples would concern such things as whether the listener were a child or an adult, a language competent or a language learner, etc.

The above variables described in three categories have been suggested in hopes that future research on the utility of visual cues to comprehension may attempt to account for greater amounts of variance in comprehension than can be accounted for at present. Before moving on to a formal model of the research area considered presently, several issues presented at the beginning of the present text will be discussed further in light of the findings reported here

The nature-nurture theme was examined in the introduction section with respect to the communication of emotions with facial expressions, where both the biological and social factors seem necessary to explain competent adult use of such behavior. It was then suggested that based on several studies (Freedman, 1974; Kimura, 1974) the same was true of gesticulation within speech. The present data indicated that encoding and decoding abilities were commonly present in all subjects tested. Although the issue may yet be an open one, it seems likely that, as was the case with facial expressions, both aspects are needed for a complete explanation. For example, there is some evidence not discussed earlier which shows how such behaviors are learned differently in various cultures (Efron, 1941). The author demonstrated how gesticualtion varied in several dialects of American English. At the same time, Condon's description of infant body-movement-reflexes to the rhythm of adult speech may indicate a biological

preparation for gesticulation.

As for the notion of a nonverbal grammar, it must be noted that no such grammar has been produced so far, which supports Birdwhistell's contention that nonverbal behavior is more dependent on context for its meaning than on any type of internal structure. In the special case of within-speech body movements, it seems quite plausable that the grammatical systems of the verbal component serve the additional nonverbal cues in a somewhat "piggy back" fashion where major verbal segments function as a mold for kinesic segmentation, if only crudely. Although the twelve stimulus items used presently were not analyzed to evaluate this contention, the analyses reported in earlier studies (Kendon, 1972) are consistent with the piggy back notion. While a grammar is rejected presently for nonverbal segments, this is not to say that an internal structure is completely missing, since such structure has been described by several students of the area (cf. Birdwhistell, 1970; Scheflen, 1974; Kendon, 1972). Figure 12b illustrates the type of structure which can be posited for kinesic cues accompanying such verbal segments as shown in Figure 12a.

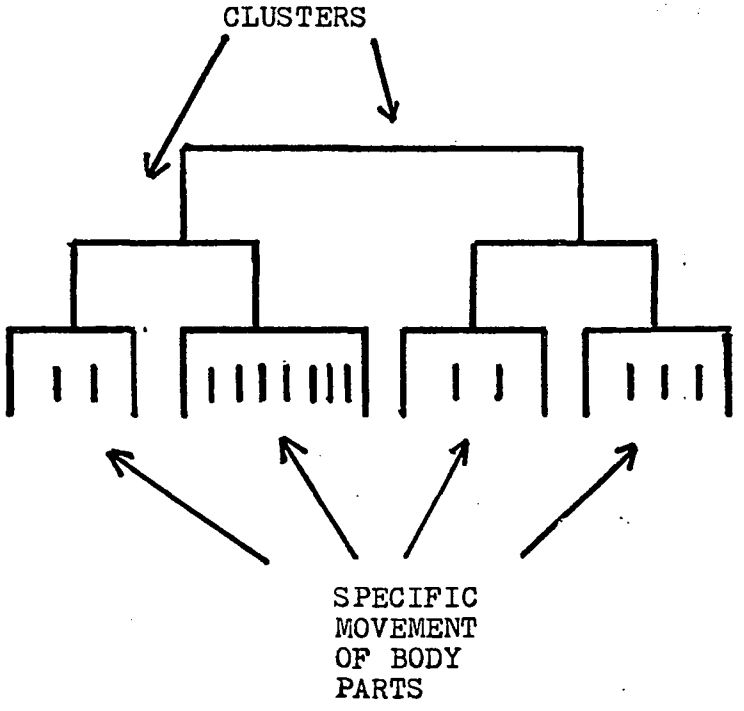
What makes the structure shown in Figure 12b different from a grammar is: (1) changes in the sequence of specific body parts do not systematically produce changes in meaning, especially since such temporal permutations may not be naturally occuring phenomena; (2) relations between clusters, or larger segments do not systematically produce higher order

Figure Caption

Figure 12. Verbal and nonverbal segments of a message (part a) and a posited nonverbal structure (part b).

VERBAL SEGMENT	NONVERBAL SEGMENT
I really don't...	head sweeps; hands open
Know why he...	eyebrow lifts; hands rotate
Would do...	head cocks; eyes open wide
Something like that.	head points down; eyes close

A



B

levels of meaning, as for example do phrases within clauses. That is, of the four kinesic meaning groups in the sample shown in Figure 12a no one segment depends on the existence of another segment for it to mean what it does. Nor is one segment more important than another nor ultimately necessary in the way that a subject, for example, is a necessary segment of a clause.

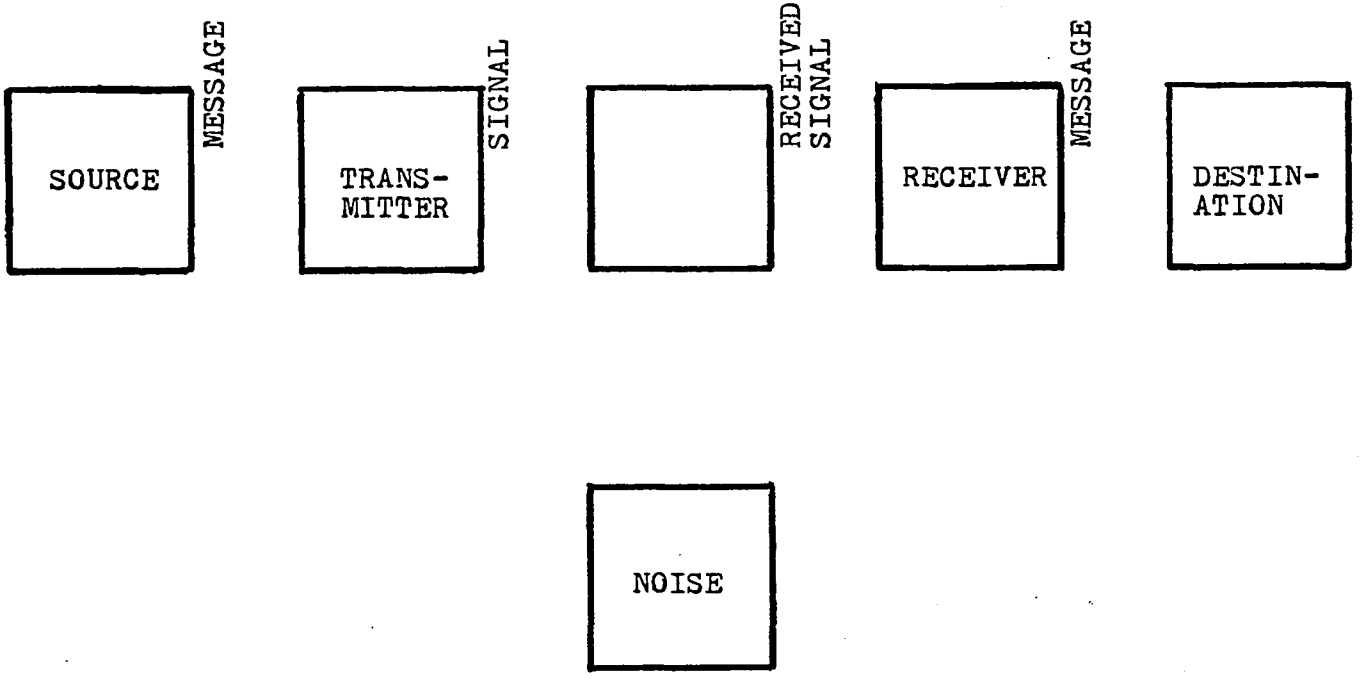
With the foregoing discussion in mind, let us turn toward the problems of developing a suitable model of the research domain considered presently. According to Deutsch (1976) a communication model (or any scientific model for that matter) should serve four functions: (1) it should organize the various elements and processes of the phenomenon; (2) it should generate ideas, heuristically, about the topic; (3) it should enable the user to make predictions; and (4) it should enable the user to measure aspects of the phenomenon in question. The widely known Shannon and Weaver model of communication (1949) demonstrates how a model can perform the above four functions, if with different degrees of success (for a resourceful user) and is shown in Figure 13.

Of course, the example-model shown above would not be adopted for the present purpose since it does not deal with many of the important variables discussed so far. The model which will now be proposed will be unfolded in three steps to facilitate explanation. The first step will incorperate a number of the observable aspects of face to face multi-

Figure Caption

Figure 13. Shannon and Weaver model (1949). Included to demonstrate how a model might organize, be heuristic, suggest predictions, and enable the making of measurements.

SHANNON WEAVER MODEL OF COMMUNICATION



modal communication:

sender receiver, context,
channel, noise,
message.

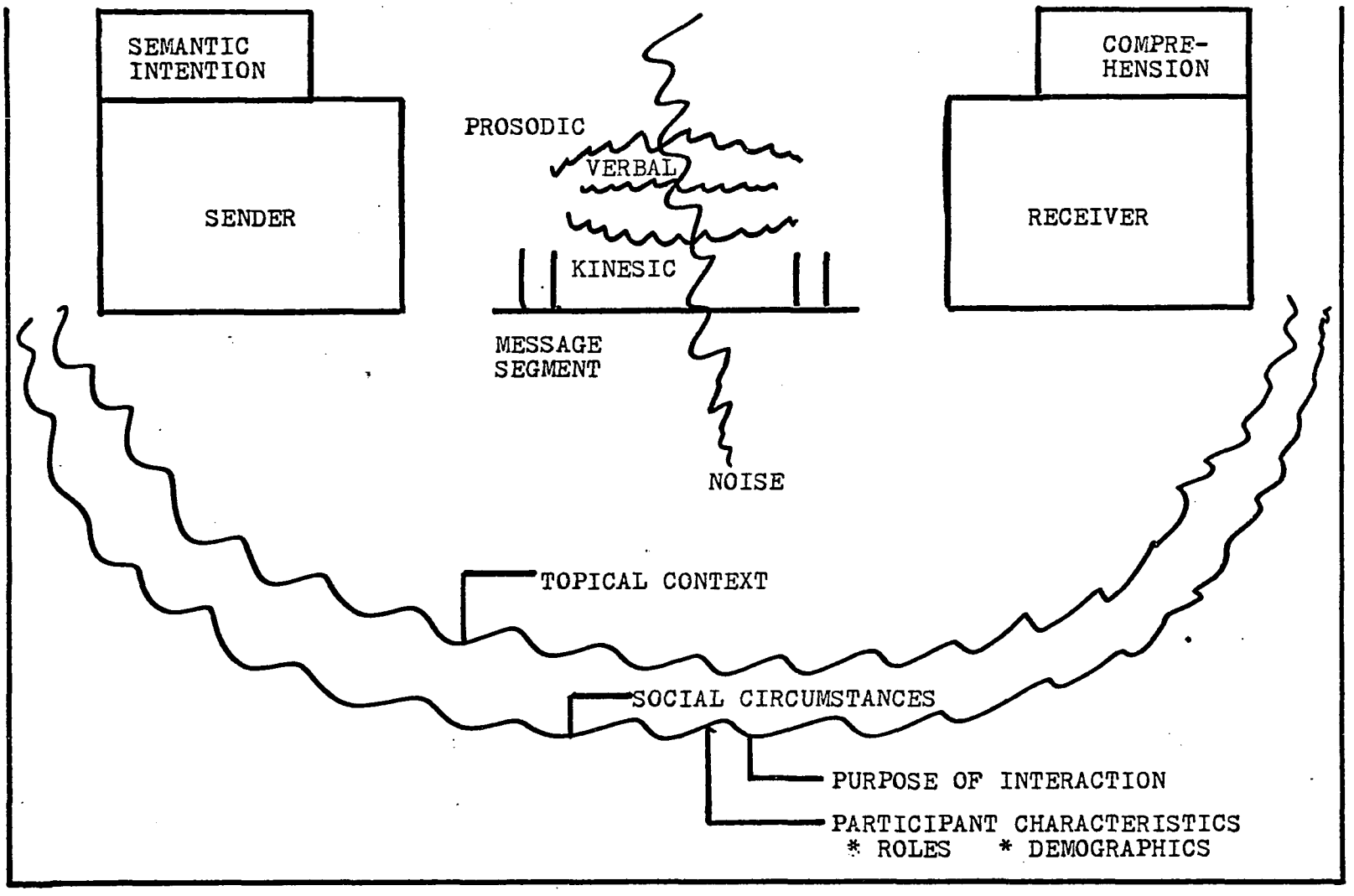
These first variables are shown in Figure 14 below.

It can be seen in Figure 14 that as a speaker encodes a message he uses at least three channels (verbal, prosodic kinesic). This event occurs in some topical and social context, while the message received will depend on both the message sent and on any noise in the system. This first approximation of a model could be used to demonstrate some aspects of the present data collection, where, for example, as the stimulus strength of the acoustic noise increases, the relative strengths of the channels change. In addition, it could be used to show how the above noise pattern can be influenced by varying the context, which would either increase or decrease the general effect noise has on comprehension.

Similarly, the model shown above can be used to relate some of the previous research cited earlier in the text. For example, the observation by Condon (1966) on the synchrony between verbal and nonverbal elements in utterances is indicated by the oscillating curves in the message area of the model. In viewing Condon's ideas visually, one could easily suspect that degree of synchrony can itself be a factor in comprehension, with extent of dysnchrony being a noise factor. In direct parallel, Dittmann's observation

Figure Caption

Figure 4. Some variables which account for multimodal speech communication.

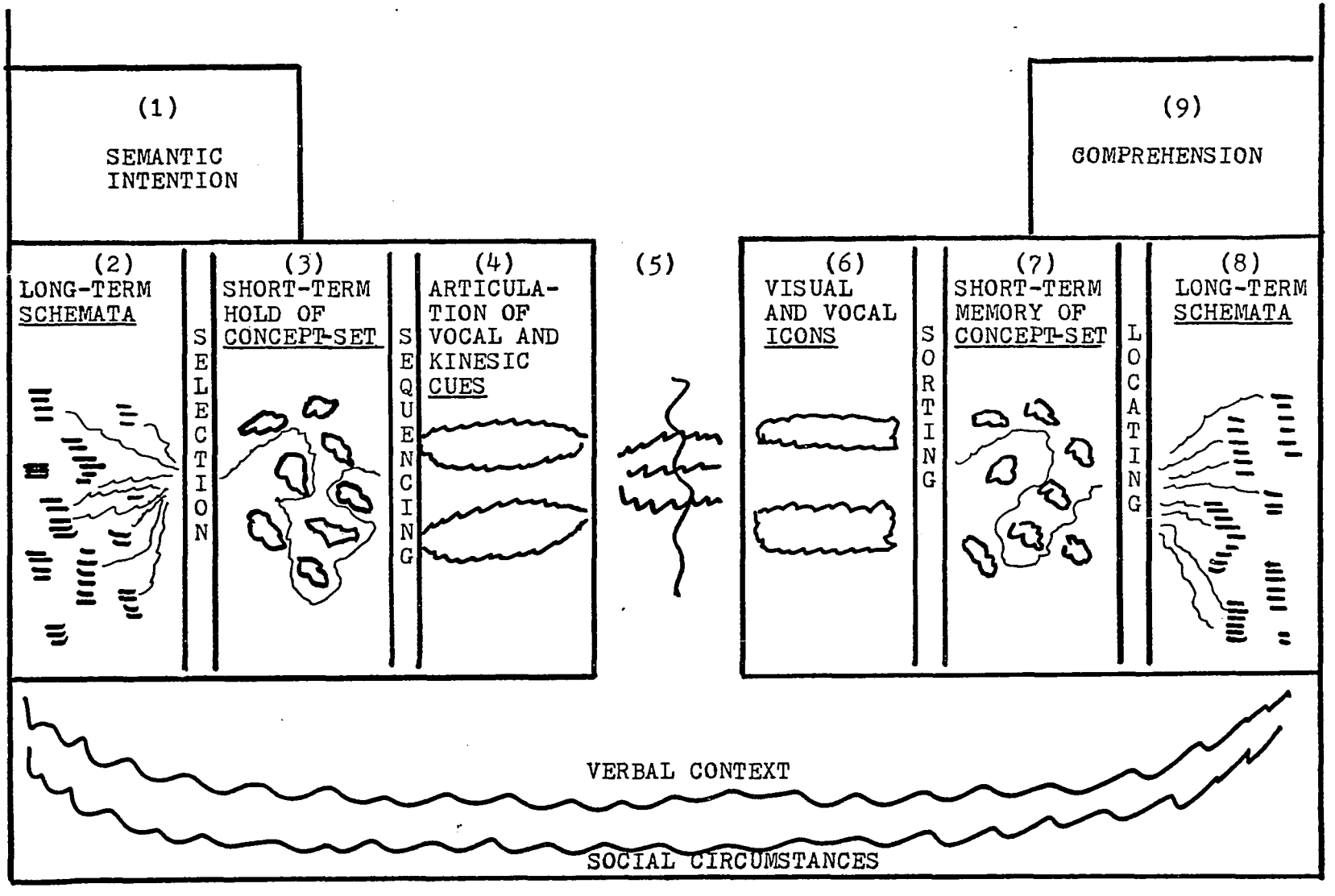

SEMANTIC INTENTION
SENDER
PROSODIC
VERBAL
KINESIC
MESSAGE SEGMENT
NOISE
COMPRE-HENSION
RECEIVER
TOPICAL CONTEXT
SOCIAL CIRCUMSTANCES
PURPOSE OF INTERACTION
PARTICIPANT CHARACTERISTICS
* ROLES * DEMOGRAPHICS

(1972) that kinesic behavior is most in evidence at the beginning of the utterance and culminates its activity at the high-information verbal-segment can be shown by permitting the oscillations to visually reflect this pattern. Here again, one might make predictions about exactly where in the utterance the kinesic utility will be at its highest. Finally, in Freedman's work (1974) on the absence of speech facilitating gestures in the utterance encoding of the congenitally blind, the cyclical nature (decoding and then encoding) of acquiring within-speech kinesic behavior is evident. One might go on to predict that within sighted populations, the extent that adult speakers gesticulate within speech will influence the extent to which such behaviors are acquired by children.

Although the model, as it has been described so far, does provide a visual framework for the present research domain, it should be pointed out that it does not reflect any of the cognitive processes going on within either the speaker or listener. Thus, several additional variables will be included at this point concerning encoding and decoding routines. According to such cognitive psychologists as Neisser (1967) and Lindsay and Norman (1972) the information processing underlying perception (and communication) involves stages whereby internal events become approximately aligned with external events through several steps. Figure 15 shows how the cognitive aspect may be included within the model being presently developed.

Figure Caption

Figure 15. Internal cognitive variables needed to account for the speech process in the face-to-face situation. Variables explained previously are shown but not labeled. The three-step cognitive process is depicted as being mirrored, respectively, in the speaker and listener segments.


(1)
SEMANTIC INTENTION
(9)
COMPREHENSION
(2) LONG-TERM SCHEMATA
SELECTION
(3) SHORT-TERM HOLD OF CONCEPT-SET
SEQUENCING
(4) ARTICULATION OF VOCAL AND KINESIC CUES
(5)
(6) VISUAL AND VOCAL ICONS
SORTING
(7) SHORT-TERM MEMORY OF CONCEPT-SET
LOCATING
(8) LONG-TERM SCHEMATA
VERBAL CONTEXT
SOCIAL CIRCUMSTANCES

What is shown in Figure 15 is a flow diagram which begins in the sender's mind as a semantic intention (1) which calls up an appropriate set of schemata out of long term memory (2). As the encoding process begins some device probably existing in short term memory holds the related set of schemata in some position (3) long enough for the sensory encoding function (articulatory movements) to execute a routine for one complete thought of approximately clause length (4). At this point, message behavior is publically available to (5) potential receivers. On the receiver's side, sensory stimuli are first temporarily stored in rich detail in the form of an icon, or brief sensory impression (6). For a relatively longer period of time analytical syntheses of the icon are held (7) for the schemata in long term memory to assimilate meaning (8). The extent to which this assimilation corresponds correctly to events external to the receiver is then called comprehension (9). The present model depicts the above described processes both in terms of vocal and kinesic phenomena.

What the additional variables do that the original set did not, is that they suugest a view of the internal behavior that enables two persons to synchronize respective, related sets of schemata by means of externally displayed behaviors. In addition, the revised model provides a framework for the positing of reasons for individual differences in both cross modal encoding and decoding skills. For example, better encoders' schemata grids may vary from

poorer ones (as suggested earlier). Again, the nurture-nature theme discussed earlier is visually illustrated by the model since the internal cognitive divisions depicted in both the sender and receiver must be biologically posited for the model to work, and since the cyclical nature of the model could be used to demonstrate how learning comes to influence cognitive type communicative functioning (either in a Piagetian sense involving the development of cognitive schemata through the processes of assimilation and accomodation, or in a Neisserian sense involving the development of cognitive schemata through the process of analysis by synthesis).

In a different way, the revised model emphasizes the limitations of the encoding and decoding task abilities at a cognitive level. That is, the processes are depicted as limited by the constraints on a short term memory function. Such constraints probably help determine a basic unit of communication in terms of what has variously been called the phonemic clause (Halliday, 1963), the prosodic phrase (Kendon, 1972), the syntagma (McNeill, 1975), or simply the common clause. In all of these terms exists the notion of one simple thought encoded in an utterance unit having one primary stress and one terminal juncture. And finally, the model enables the user to consider cognitive dysfunction in relation to the communicative process. For example, DeRenzi et al (1968) referred to those persons with both verbal and gestural disabilities as being conceptually disordered. This condition is clearly representable by the model whereby disorders may

be understood to exist at various levels of behavioral abstraction (i.e., higher or lower cognitive processes) corresponding to disturbances within various sections of the model's representation of speaker and listener activities (Brown, 1972).

One last modification of the model will be briefly considered at this point. That is, all statements in a conversation (except the first) occur in the context of one or more previous statements which can readily affect how the given statement is intended and understood. Very likely this factor is of greater importance than is usually conceded. This last variable is shown in Figure 16, which represents the completed form of the model suggested presently. With the addition of this last variable, one might speculate that comprehension accuracy is related to the number of elapsed prior statements in a conversation or that the overall extent of gestural accompaniment might vary relative to the number of elapsed statements (since information on a given topic increases as the conversation continues). For example, the phrase "move in" generally indicates a type of motion. However, in the context of a robbery-group discussing a prospective bank-holdup, the phrase might well signify a motion involving: swiftness, stealth, potential agression, purpose, and on-the-spot reconnaissance. One might even predict that the information value of a particular gesture, as well, would vary in relation to the amount of previous conversational context.

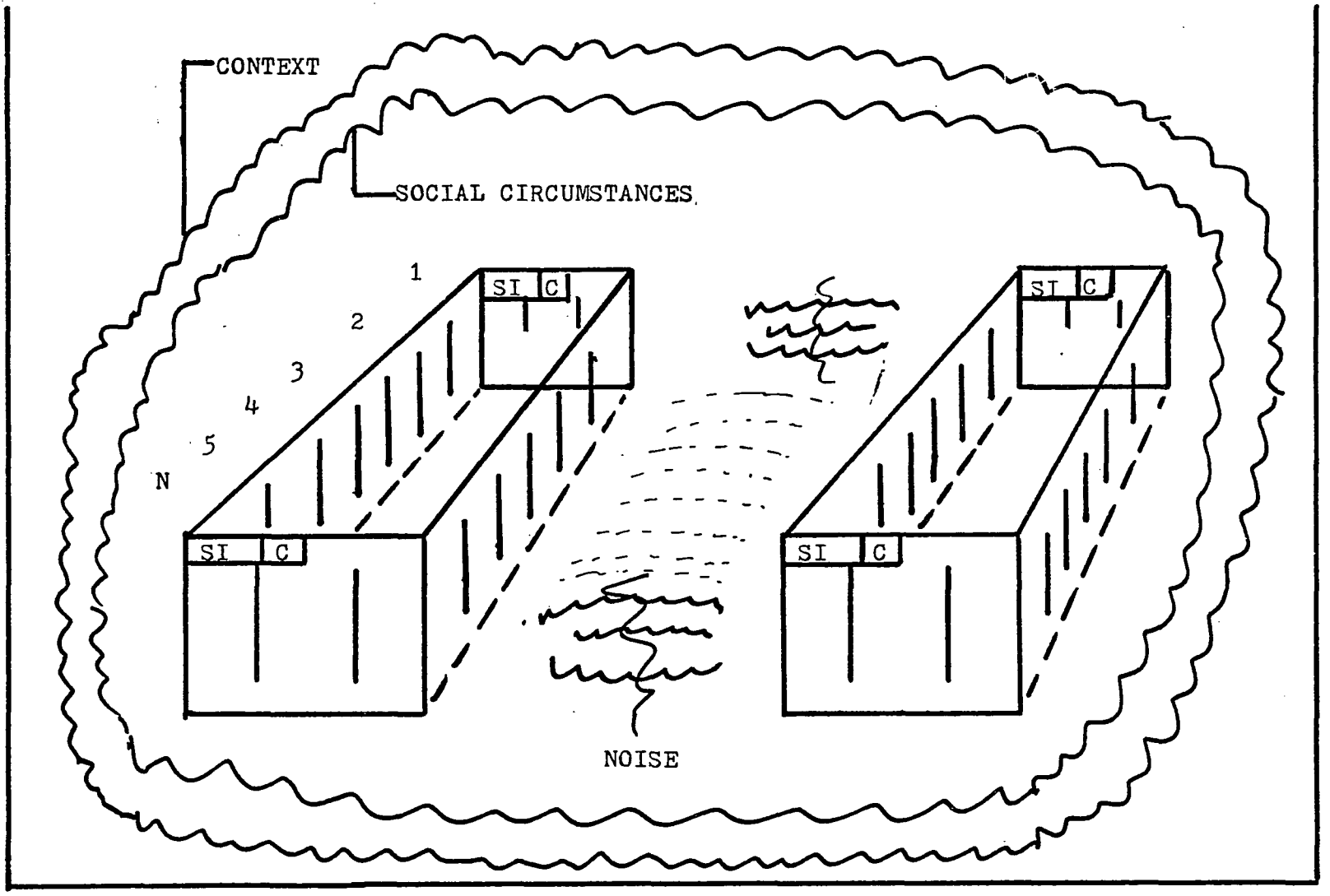
In the above illustration, for example, the gang leader might have said, "and then we pull the car into the alley, cut the alarm, and then we..." / thumb jerks in a lateral direction /.

Hopefully the foregoing description of a multi-modal model has made two things possible. First, it is hoped that the model will generate questions about the two experiments conducted presently that were not evident prior to the development of the model. And second, it is hoped that a set of possible future studies will be more efficiently generated now with it, than would have been the case without it.

As for the first objective, let us return for the moment to the version of the model shown in Figure 15. In the earlier discussion of why listeners did better when they could see the speaker than when they could not, it was suggested that perhaps their attention was better maintained with the addition of visual cues or that perhaps bimodal signal redundancy was responsible. In viewing the receiver part of the model, on the right side of Figure 15, it can now be seen that the visual utility factor might be accounted for in any of four locations shown here. That is, it might occur because of increased, or better focused, attention (6), or because of a richer bimodal sensory image or icon (7); likewise, it might occur because of some interaction effect between short term visual and vocal analyses-syntheses (8), or because of a richer bi-

Figure Caption

Figure 16. The completed model, including the variable of conversational context. The upper part of the diagram indicates the first communicative transaction in the conversation. The lower part indicates the nth communicative transaction. Thus, all previous transactions serve as a context for the speaker's and listener's communicative processing. Both a semantic-intention area and a comprehension area are shown in the two parties, since each in reality will both encode and decode.


CONTEXT
SOCIAL CIRCUMSTANCES
1
2
3
4
5
N
SI
C
SI
C
SI
C
SI
C
NOISE

modal excitation of long term held schemata. Thus, it seems the model does suggest explanations in a systematic way.

Similarly, with respect to the speaker side of the model in Figure 14, it was earlier suggested that individual differences in speakers' kinesic outputs (and linguistic outputs for that matter) may be qualitatively different in communicative utility depending on such things as cognitive complexity, intelligence, etc. The model suggests, however, several other ways that utility of kinesic cues may be affected. The most obvious concerns the schemata area where content is first excited. Certainly, the degree to which this content is codable either verbally or nonverbally will influence how much kinesic encoding will take place. Obviously, utility will increase if either the material is especially codable kinesically or especially not codable verbally.

In a very different way, the model provides a view of the interaction between the cognitive and emotive dimensions of communication. Although ideas can be communicated either verbally or nonverbally and although emotions can be communicated verbally or nonverbally, a measure of specialization does exist, with the kinesic channel better serving the communication of emotions and the verbal channel better serving the communication of ideas. But of course, the overlap area, where for example, gestures can assist in the communication of ideas, is indeed the topic presently addressed. It is just possible, however, that increasing activity in

emotional output through nonverbal behavior during moments of verbalization may bring about a type of "spill-over" effect where this increased activity in the kinesic channel (for emotive reasons) will in addition increase the utility of such cues for the communication of the cognitive information which is simultaneously being emitted in the verbal channel.

In the foregoing discussion, several contentions about the functions of kinesic behavior with the speech process have been made which seem possible, based on the limited research conducted in the area so far, but which have yet to be supported by any direct research. In the following pages several possible studies will be suggested which might further the limited knowledge we have at this point on the utility of kinesic cues in the speech process. The general topic areas such future studies would fall into are: (1) normal adult processes, (2) child acquisition patterns, (3) cross cultural descriptions, and (4) pathologies. One or two brief suggestions will be made for each of the four above areas.

Normal adult processes. It was promised earlier that a revised strategy for studying the utility of kinesic cues in terms of the communication of speech acts would be forthcoming. To alleviate the severe difficulty normal subjects have in viewing speakers without the presence of sound, the following could be done. Using trained actors, instances of particular speech acts (requests, promises, demands,

negations, warnings, etc.) would be assembled on a presentation tape. Actors would be instructed to encode speech acts as realistically as possible. The following procedure should serve as an example of how testing would be conducted.

One of the test items might be:

1. John saves his money habitually.

The actors would be requested to utter this sentence with each of three of the following impacts (illocutionary forces) --affirmative assertion, questioning of the truth value, and negation (as when said as a retort, meaning "you must be crazy if you think that"). In the actual test, subjects would first be presented with a written transcript of the sentence, after which they would view the visual portion of utterances, as done previously. Subjects would respond to multiple choice items corresponding to the various speech act treatments given to individual sentences by the actors. Also, since certain answers might seem more likely to subjects than others, all sentence treatments would need to be used with different subjects within a balanced presentation schedule.

In addition to the above possibility, a number of follow up studies might be performed, all of which could seek to further clarify the general degree of utility found in the present study for kinesic cues in relation to speech comprehension. For example, <u>context</u> could be varied whereby the procedure used in Experiment 1 would in addition control the amount of prior conversational statements subjects would hear

and see before they received the stimulus statement. It is expected that the utility of visual cues would increase as the subjects' contextual knowledge increased. Content might also be varied by generating stimuli from a more varied set of semantic domains than was used presently in order to assess in specific terms how utility exploits or does not exploit kinds of semantic meaning. For example, in addition to spatial and abstract topics (economics, education, etc.) topics concerning human relations, or literature might also be employed. Finally, speakers with various personalities or speakers under various degrees of emotional arousal might be studied to evaluate the roles of these two variables with respect to kinesic utility. The above study proposals could all be carried out with the same general design used in the present study.

Cross cultural descriptions. The nature-nurture theme discussed earlier is relevant to this topic area because the extent to which different cultures do or do not utilize kinesic support behaviors in a similar manner reflects on how universal (or culture-bound) such behaviors are in their ontogeny. One approach here would be to give individual speakers in different language cultures the same semantic content to express in their repective languages, provided such content was relatively equally verbally codable. It would then be interesting to observe how similarly or dissimilarly speakers in different language groups would mark particular semantic aspects of given propositions. Such observations would concern: (1) whether particular propositional items

are kinesically marked or not; (2) in what particular form they are marked; (3) when and where they are marked; and (4) how much utility to speech comprehension results from various manners of marking. Additionally, hypotheses concerning kinesic support-form similarity might be couched in terms of language-family memberships.

A second approach to examining cross cultural relationships on the present topic might be to test listeners who speak one language about what speakers of a foreign language are saying. This examination could be accomplished by varying kinesic cue availability, similar to the way done presently, and then testing semantic comprehension. A multiple-choice format would be used, keeping in mind the general difficulty of the subject's task when drafting questions. The extent that listeners are able to utilize body movements to improve comprehension scores will provide evidence to support a universal theory of kinesic utility, if only in a limited way. That is, to the extent that such cues are iconically coded they may be considered universal in use while to the extent that they are coded arbitrarily they may be considered culture specific.

Child acquisition patterns. In earlier discussion it was speculated that the quality and quantity of parental within-utterance kinesic behavior ought to play some role in certain aspects of the child's acquisition of communicative competence. Of course some objective measure of "quality" of output must exist before attempting to carry out a study

based on this contention. Four reasonable categories are suggested here:

(1) An index of the incidence of kinesic semantic markers (number of such markers divided by the number of elapsed words in a given speech sample (token measure); number of different semantic markers divided by the number of elapsed words (type measure).

(2) An index of the incidence of kinesic punctuation markers (number of tokens divided by the size of sample; number of types divided by the size of sample).

(3) An index of the degree of synchrony or fluidity of the speech performance (high, moderate or low; this observation would require training before judgements could be made).

(4) An index of body-part utilization (number of different body parts used in some standard sized speech sample (50 or 100 words, etc,); sample size must be a constant since the numerator of index is limited while the denominator is unlimited).

The total scoere would equal the sum of indicies in categories 1 through 4 (with category 3 converted into some numerical value which would give it equal weight relative to the other three categories. Standardized comparisons would obviously be useful here.

With the above measuring instrument in hand, parent-child dyads could be studied by evaluating the "quality of parental output and then observing the respective child's

abilities on some set of communication skills. The most obvious observation would be to evaluate the child's within-speech kinesic output, since this would enable some assessment of to what extent (and when) children copy this type of behavior from their parents.

A second type of developmental observation which could be made concerns describing the utility children make of kinesic cues as listeners at various ages (one study reviewed earlier (Miahael & Willis, 1968) did in fact study the developmental use by children of a limited set of highly stylized gestures such as converying the meaning of "hi," "ok," and "crazy." All of these gesture types can be used out of a speech context). A standardized comprehension test similar to the one used in the present study ought to be constructed so as to include at least five variables within its testing sensitivity:

(1) Quality of the speaker's kinesic output would vary, possibly in terms of the four categories described earlier. Specific test stimuli would be categorized as "high" or "low" quality of kinesic output.

(2) The content of the aural-visual stimuli would vary. Examples of different type content would be: abstract topics such as about wealth, education, and politics; spatio-temporal descriptions of moving objects; human relations problems, such as why someone is angry; simple scientific descriptions, such as about water displacement or about why objects fall when they are dropped;

statements about a story, such as "Little Red Riding Hood" or "The Three Bears."

(3) Signal to noise ratio would vary simply in terms of either easy-to-hear or hard-to-hear circumstances.

(4) Context would vary. This refers to the amount of information given to the listener about the topical or conversational contexts a given stimulus was drawn from. Such information would be present or not present.

(5) Channel would vary in terms of whether visual cues were present or not present.

To facilitate the use of such a test with young children, the test would have to be administered orally.

Data collected with the above test could be used to evaluate the role of kinesic cues within speech processing in terms of several developmental issues, including: (1) the importance of the adult's kinesic fluency (within speech) to the child's overall ability to understand messages; (2) how children's ability to decode kinesic acts within various semantic domains can be compared with related literature on cognitive development; and (3) the role noise and contextual information paly in childrens use of kinesic cues to decode messages. Certainly the well known theories of cognitive development (e.g., Piaget) could be used to generate hypotheses for the above proposals. For example, younger children might be expected to be more dependent on kinesic communication of concepts in face-to-face sending-receiving situations since they conceptualize at a more

concrete level (as opposed to symbolic level). McNeill (1975) makes this very contention with respect to the ontological development of adult-gesture out of a sensory-motor intelligence.

Pathologies. While, of course, several of the proposals discussed so far might be extended for study of populations with various kinds of communication problems, two additional possibilities in particular will be proposed. First, an additional diagnostic tool might be developed in terms of the measurement of synchrony between a speaker's verbal and kinesic behavior. Condon (1966) made this speculation based on his study of behavioral synchrony in terms of certain neurological impairments observable in terms of the degree of dysynchrony. Thus, with the encoding-skills observation scheme discussed earlier and with some objective measure of degree of synchrony, a trained observer might be able to monitor in an aural-visual way where and when the encoding process were faltering or recuperating at specific points within utterances. Such a motoric index might prove to provide an additional diagnostic service to clinicians.

A second possibility must be stated in very general terms due to the present writer's very limited capacities in the area. However, at several times in the present text a close cognitive relationship seemed a reasonable posit with respect to the encoding of internal schemas through both verbal and kinesic cues. It is suggested here that students of clinical methods of speech therapy (or communication

therapy) might investigate the possibility that some type of "gesture training" could be facilitative to the overall rehabilitation of persons with communicative pathologies. In either the encoding or decoding communicative activity, the assumption is that stimulation of nonverbal processing of internal concepts might improve verbal processing of internal concepts, especially if both were dealt with in conjunction.

Footnotes

[1]Representative bibliographies concerning all of the above studies appear in Birdwhistell (1970), Eisenberg and Smith (1971), and Knapp (1972).

[2]Admittedly, the same is true of a word or any larger verbal segment. However, the extent to which kinesic acts seem dependent on context represents a significant leap over the extent to which verbal segments are context dependent.

[3]The term "speech act" refers to the formulation by Austin (1962), and later refinement by Searle (1969), that verbal behavior operates on at least three levels of analysis, (1) the locutionary level, which refers to the act of pronouncing words in the form of an utterance; (2) the illocutionary level, which refers to the act of performing social deeds with an utterance, such as asking a question, making a promise, or giving a command; and (3) the perlocutionary level, which concerns the effects a given speech act may have on a listener, such as persuading, frightening, or assuring.

[4]Such a possible design will be suggested in the general discussion section of the present text.

[5]For Experiment 1 lip cues and kinesic cues were not formally differentiated, as Experiment 2 concerns this topic. However, an indirect assessment is made in the first experiment on the effects of both types of cues on speech comprehension.

[6]Reasons for the different effects produced by the two groups of items will be discussed later, in the results section.

[7]Lip cues are simply "fall-out" signs directly dependent on verbal behavior. In a sense, they represent an impovershed verbal system in a visual mode. Kinesic cues, on the other hand, are not coded in a linguistic fashion and comprise a partially iconic, partially arbitrary, separate communication system which is synchronized to various extents with utterances during moments of verbalization (this distinction does not hold for the sign behavior of the blind).

8

Since the method employed in Experiment 2 in most ways follows the method employed in Experiment 1, details which are totally redundant are not included here. The reader is advised to refer back to the description of method in Experiment 1 for complete details.

APPENDIX A

TEST OF COMPREHENSION AND MEMORY OF INFORMATION

Instructions:

The following is a test of your ability to understand and remember information.

You will HEAR one statement at a time. There are 12 statements in all. After hearing each statement, you will be asked to answer questions in order to test how much you have understood and remembered from each of the statements.

Do not take a long time to answer any particular question. UNDER NO CIRCUMSTANCES should you go back and change a previous answer (or answer it if you had not done so before) once you have read the next question or questions.

Before the ACTUAL statements and questions are presented to you, several SAMPLE statements and questions will be given to you, in order to make sure that you are familiar with what you have to do.

Turn to the next page (after filling out the bottom of this page) and stop there, until you are requested to proceed further.

Name _____ Age _____

Sex _____ Date _____

Form _____

Questions for First
Tape Sample

1. One of the main things described in the statement was:
   a. a television
   b. a radio
   c. a record player
   d. a game
   e. a crane
   f. a bird

2. When the thing described above is ready to begin its action, some specific part of it is located:
   a. at the top of something
   b. at the bottom of something
   c. inside of something
   d. alongside of something
   e. under something
   f. not near something

3. Next, another specific part of the thing described above goes:
   a. inside out
   b. out to the side
   c. over a bump
   d. around a pole
   e. under a cover
   f. away completely

4. Next, something:
   a. jumps up
   b. drops down
   c. moves away
   d. goes over a bump
   e. goes away from a large hook
   f. slowly moves up

5. Finally, one particular part of the thing:
   a. completely stops
   b. falls away
   c. goes to the beginning
   d. disappears
   e. gets louder
   f. gets softer

Questions for Second
Tape Sample

1. One of the main things described in the statement was:

   a. a basketball court
   b. a hardware store
   c. an equipment room
   d. a shipment room
   e. a waiting room
   f. a railroad station

2. One aspect of the thing described in the statement is that:

   a. the speaker is sitting in it
   b. it is on fire
   c. it is empty
   d. the speaker just left it
   e. the speaker owns it
   f. people are playing in it

3. Another aspect of the thing described in the statement is that:

   a. it is a center for something
   b. it is used for fun and exercise
   c. its were you can buy something
   d. it is off limits to certain people
   e. it is for sending things
   f. it serves no particular purpose

4. Something is kept there before it is used:

   a. trains
   b. hardware
   c. boxes
   d. people
   e. sports equipment
   f. stamps

APPENDIX B

1. The main thing that did something in the statement was a(n):
   a. man
   b. woman
   c. cabdriver
   d. foreman
   e. parking lot attendant
   f. boy

2. One action that the thing above did was that it:
   a. hit something
   b. skidded on something
   c. picked-up something
   d. rolled something
   e. drove something
   f. followed something

3. Another action that the thing above did was that it:
   a. speeded up a lot
   b. reversed direction
   c. slowed down a little
   d. started abruptly
   e. did not alter its action
   f. went around and around

4. In doing what it did, the main thing in the statement at one point went:
   a. down
   b. over
   c. to the right
   d. under
   e. to the left
   f. in between

5. Finally, the main thing ended up going:
   a. in the opposite direction
   b. in a different direction
   c. in the wrong direction
   d. in the same direction
   e. in no direction
   f. in the correct direction

1. One of the main things described in the statement was:

   a. liberation
   b. laws
   c. courts
   d. civilians
   e. objections
   f. necessities

2. The main thing described above was said to be:

   a. right
   b. not important
   c. wrong
   d. unnecessary
   e. necessary
   f. antisocial

3. One purpose of the thing described in the statement was that it:

   a. liberates people
   b. protects people
   c. civilizes people
   d. encourages people
   e. assists people
   f. educates people

4. Another purpose of the thing described in the statement was that it:

   a. would free something
   b. would popularize something
   c. would populate something
   d. would regulate something
   e. would hold back something
   f. would annihilate something

5. The "something" described in the previous question is:

   a. individuals
   b. criminals
   c. society
   d. rights
   e. consumers
   f. property

1. One of the main things that did something in the statement was:

   a. two owls
   b. three owls
   c. many owls
   d. two bowls
   e. three bowls
   f. many bowls

2. Air was doing something to the main things in the statement. Specifically, air was:

   a. circulating around
   b. blowing from left to right
   c. blowing from right to left
   d. ventilating down
   e. ventilating up
   f. not moving in any one direction

3. One thing that was happening to the main things in the statement was that it was:

   a. moving up
   b. remaining motionless
   c. going around
   d. breaking into pieces
   e. being stopped
   f. moving down

4. What was happening to the main things in the previous question was happening in relation to:

   a. the floor
   b. the ceiling
   c. a corner
   d. each wall
   e. a fly
   f. each other

5. Finally, what was happening to the main things was said at the end of the statement to NOT concern the general aspect of:

   a. left to right
   b. up and down
   c. in and out
   d. over and under
   e. around and around
   f. with and against

1. One of the main things described in the statement was:

a. gold
(b) money
c. credit
d. checks
e. tax
f. jewels

2. The main thing described above was said to be:

a. an exchange medium
b. a change medium
c. an interchange medium
d. an exchange of compensation
(e) a change of compensation
f. an interchange of compensation

3. The main thing described in the statement was said to be:

a. unused
b. sold
c. unsold
d. reasonable
(e) used
f. abused

4. One purpose of the main thing in the statement was said to concern:

a. consolation
b. compensation
c. cooperation
d. reason of mind
(e) debt
f. credit

5. Something made possible by the main thing in the statement concerns:

a. the acquisition of services
b. the acquisition of goods and services
c. the sale of goods
(d) the sale of goods and services
e. the functioning of government services
f. the functioning of good government services

1. One of the main things that is described in the statement was a:

a. man
b. boy
c. painting
(d.) film
e. sign
f. window

2. One action that the main thing described above is involved with concerns bringing our attention to a:

a. freak
b. woman
(c.) image
d. pigeon
e. chicken
f. pig

3. When the thing above is observed it seems to be:

a. wavering
b. weaving
(c.) resting
d. yawning
e. screaming
f. disappearing

4. Also, the thing above is said to be doing something in relation to:

(a.) water
(b.) grass
c. glass
d. mud
e. fodder
f. food

5. Also, the thing above is said to be watching:

a. some of the time
b. at no time
(c.) all of the time
d. carefully
e. curiously
f. cautiously

1. The main thing that did something in the statement was:

a. a hall
b. a pebble
c. a nut
(d.) a ball
e. a boxer
f. a little kid

2. The main thing was said to be:

a. heavy
b. light
c. sandy
(d.) quick
e. green
f. brown

3. The main thing in the statement at one point was said to be:

(a.) in a corner
b. in a box
c. on some grass
d. under a cover
e. in a bottle
f. over a board

4. At the end of the statement, the thing:

a. came off a wall
b. came onto a table
c. came off the ceiling
d. slid off a rock
e. broke into pieces
(f.) bounced straight up

5. As the main thing does its last action, it does it:

a. directly
b. indirectly
c. slowly
(d.) quickly
e. continuously
f. only sometimes

1. One of the main things described in the statement concerns:

   a. economics
   b. elocution
   c. education
   d. execution
   e. energy
   f. eternity

2. The purpose of the main thing in the statement concerns:

   a. society
   b. sincerity
   c. speech
   d. art
   e. time
   f. music

3. One of the things that can happen to the main thing is that it can be:

   a. forgotten about
   b. felt for
   c. evened out
   d. passed out
   e. looked over
   f. passed on

4. What happens to the main thing specifically can concern:

   a. religion
   b. literature
   c. licenses
   d. movements
   e. money
   f. doctrine

5. Also, what happens to the main thing can concern:

   a. some other forms
   b. all other forms
   c. no other forms
   d. some other facts
   e. all other facts
   f. no other facts

1. One of the main things described in the statement was a:

   a. bus
   b. rail car
   c. elevator
   d. boat
   e. subway
   f. monorail

2. One action that the main thing was doing concerned its proceeding:

   a. through an opening
   b. along a path
   c. around a turn
   d. over a bridge
   e. under a tunnel
   f. within an enclosed area

3. One aspect of the above action concerned:

   a. going down at an angle
   b. turning right
   c. going straight up
   d. turning left
   e. going up at an angle
   f. going straight down

4. As the main thing did what it did, it did so by:

   a. moving steadily
   b. moving only some of the time
   c. remaining in one place
   d. moving abruptly
   e. going in circles
   f. appearing slowly

5. The main thing finished what it was doing:

   a. on the ground floor
   b. at an intersection
   c. in a station
   d. in a parking lot
   e. in the water
   f. where it started

1. The main thing that did something in the statement was a:

a. moving vehicle
b. moving horse
c. boy walking
d. parked vehicle
e. horse standing still
f. boy standing

2. As the main thing does what it does, ~~xxxxxxxxxxxx~~ it (he) encounters a:

a. policeman
b. dark object
c. set of lights
d. fence
e. forked path
f. one way street

3. At one point, the main thing does something by using:

a. a film
b. a wheel
c. a light
d. two films
e. two wheels
f. two lights

4. At another point, the main thing acts by:

a. running over to something
b. sneaking over to something
c. inching over to something
d. going directly over to something
e. avoiding going over to something
f. sliding over to something

5. The "something" refered to in the previous question was:

a. the other street
b. the other bush
c. the other side
d. the other section
e. the other beat
f. the other slide

1. The main things described in the statement are:
   a. tools
   b. buildings
   c. courts
   d. pools
   e. governments
   f. utensils

2. The main thing in the statement is described as:
   a. being something
   b. not being something
   c. having something
   d. not having something
   e. costing something
   f. not costing something

3. The main thing in the statement is described as being:
   a. little
   b. minor
   c. active
   d. large
   e. vital
   f. idle

4. The main thing is described as:
   a. becoming necessary in the future
   b. becoming unnecessary in the future
   c. being necessary now
   d. being unnecessary now
   e. having been necessary in the past
   f. having been unnecessary in the past

5. The main thing is described in the statement in relation to something that is:
   a. theirs
   b. no ones
   c. ours
   d. everyones
   e. one persons
   f. several persons

1. The main thing that did something in the statement was a:

a. pot
b. river
c. sink
d. fountain
e. boiler
f. brook

2. One action that the main thing did concerned:

a. crushing something
b. changing something
c. shooting out something
d. holding in something
e. dropping something
f. saving something

3. The action being done takes place:

a. on top
b. on the bottom
c. on one side
d. on two sides
e. in the middle
f. in no particular place

4. The "something" that is the object of the above action is being:

a. collected
b. dissected
c. directed
d. scattered
e. splattered
f. contained

5. The above object of the main action is being acted upon:

a. cautiously
b. readily
c. sloppily
d. haphazardly
e. slowly
f. rapidly

1. The main thing that was described in the statement was a:

   a. video telephone
   b. recorder
   c. printed circuit
   (d.) projector
   e. camera
   f. sewing machine

2. One action that is described concerns something:

   a. passing something
   b. pushing something
   (c.) pulling something
   d. bypassing something
   e. putting something in place
   (f.) pulling out something

3. The "something" described in the previous question is a(n)

   a. lock
   b. link
   c. socket
   d. dial
   e. sprocket
   (f.) connection

4. Another action that is done by the main thing is that it causes something to:

   a. stop
   b. go away from something
   c. start
   (d.) connect
   e. go along with something
   f. go through something

5. Finally, ~~hmm~~ something ends up:

   (a.) in a box
   b. under a cover
   c. exactly where it began
   d. switching off
   e. in some distant location
   f. on a holding device

REFERENCES

Austin, J. How to do things with words. Fair Lawn, N. J.: Oxford University Press, 1962.

Baxter, J., & Winters, E. Gestural behavior during a brief interview as a function of cognitive variables. Journal of Personality and Social Psychology, 1968, 8, 303-307.

Birdwhistell, R. Introduction to kinesics. Louisville: University of Louisville Press, 1952.

Birdwhistell, R. Some body elements accompanying spoken American English. In A. Smith (Ed.). Communication and culture. New York: Holt, Rinehart and Winston, 1966.

Birdwhistell, R. Kinesics and context. Philadelphia: University of Pennsylvania Press, 1970.

Brown, J. Aphasia, apraxia and agnosia: clinical and theoretical aspects. Springfield, Ill.: Charles Thomas, 1972.

Campbell, D. & Stanley, J. Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Company, 1963.

Chapanis, A. Prelude to 2001: explorations in human communication. American Psychologist, 1972, 27, 949-961.

Condon, W. Method of micro-analysis of sound films of behavior. Behavior Research Method and Instrumentation, 1970, 2, 51-54.

Condon, W. & Ogston, W. Sound film analysis of normal and pathological behavior problems. Journal of Nervous

and Mental Disease, 1966, 143, 338-347.

Condon, W. & Sanders, L. Neonate movement is synchronized with adult speech: interactional participation and language acquisition. Science, 1974, 183, 99-101.

Darwin, C. The expression of emotions in men and animals Englewood Cliffs, N.J.: Appleton Centry Crofts, 1896.

Delong, A. Kinesic signals at utterance boundaries in pre-school children. Semiotica, 1974, 11, 43-73.

DeRenzi, E., Pieczuro, A. & Vignolo, L. Ideational apraxia: a quantitative study. Neuropsychologia, 1968, 6, 41-52.

Deutsch, K. Cited in J. DeVito. The interpersonal communication book. Scranton, Pa.: Harper and Row, 1976.

Dittmann, A. The body movement-speech rhythm relationahip as a cue to speech encoding. In A Siegman & B. Pope (Eds.) Studies in dyadic communication. New York: Pergamon Press, 1972, 135-151.

Dunlap, K. The role of eye-muscles and mouth muscles in the expression of the emotions. Genetic Psychology Monographs, 1927, 2, 199-233.

Efron, D. Gesture and environment. New York: Kings Crown Press, 1941.

Eisenberg, A. & Smith, R. Nonverbal communication. New York: Bobbs-Merrill, 1971.

Ekman, P. Body position, facial expression and verbal behavior during interviews. Journal of Abnormal and Social Psychology, 1964, 48, 295-301.

Exline, R. Affective relations and mutual glances in dyads. In S. Tomkins & C. Izard (Eds.) Affect, Cognition and Personality. New York: Springer, 1965, 319-330.

Fillmore, C. Types of lexical information. In D. Steinberg & L. Jakobovits (Eds.) Semantics. New York: Cambridge University Press, 1971.

Freedman, N., Ohanlon, J., Oltman, P. & Quitkin, H. The imprint of psychological differentiation on kinetic behavior in varying communicative contexts. Journal of Abnormal Psychology, 1973a, 79, 239-258.

Freedman, N., Blass, T., Rifkin, A. & Quitkin, F. Body movements and the encoding of agressive affect. Journal of Personality and Social Psychology, 1973b, 26, 73-85.

Freedman, N. & Steingart, I. Body movement and verbal encoding in the congenitally blind. Perceptual and Motor Skills, 1974, 39, 279-293.

Freud, S. Fragment of an analysis of a case of hysteria. Collected papers, 1905, 3 (New York: Basic Books, 1959).

Graham, J. & Argyle, M. A cross cultural study of the communication of extra-verbal meanings in gestures. International Journal of Psychology, 1975, 1, 21-28.

Graham, J. & Heywood, S. The effects of the elimination of hand gestures and of the verbal codability on speech performance. European Journal of Social Psychology, 1975, 3, 3-10.

Haley, J. Strategies of psychotherapy. New York: Grune and Stratton, 1963.

Halliday, M. The tones of English. Archives Linguist, 1963, 15, 1-28.

Hewes, G. Gesture language in culture contact. Sign Language Studies, 1974, 1, 1-34.

Hewes, G. Current status of the gestural theory of language origin. In origins and evolution of language and speech. Annals of the New York Academy of Science, 1976, 280, 482-503.

Johansson, G. Visual motor perception. Scientific American, 1975, , 76-87.

Kendon, A. Some relationships between body motion and speech. In A. Siegman & B. Pope (Eds.) Studies in Dyadic Communication. New York: Pergamon Press, 1972, 177-208.

Kimura, D. Manual activity during speaking--I right handers. Neuropsychologia, 1973a, 11, 51-55.

Kimura, D. Manual activity during speaking--II left handers. Neuropsychologia, 1973b, 11, 56-60.

Kimura, D. & Archibald, Y. Motor functions of the left hemisphere. Brain, 1974, 97, 337-350.

Knapp, M. Nonverbal communication in human interaction. New York: Holt, Rinehart & Winston, 1972.

Lindsay, P. & Norman, D. Human information processing. New York: Academic Press, 1972.

Lindsenfeld, J. Verbal and nonverbal elements in discourse. Semiotica, 1971, 3, 223-233.

Lindsenfeld, J. Syntactic structure and kinesic phenomena in communication events. Semiotica, 1974, 11, 38-46.

McNeill, D. Semiotic extension. Paper presented at the Graduate School of the City University of New York, April, 1975.

Mehrabian, A. Inference of attitude from posture, orientation and distance of communicator. Journal of Consulting and Clinical Psychology, 1968, 32, 296-308.

Michael, G. & Willis, F. The development of gestures as a function of social class, education, and sex. Psychological Record, 1968, 18, 515-519.

Munari, B. Supplemento as Dizionario Italiano. Milan: Muggiani Editore, 1963, 66-67.

Neisser, U. Cognitive psychology. Englewood Cliffs, N.J.: Prentice Hall, 1967.

Scheflen, A. How behavior means. New York: Anchor Press, 1974.

Schlosberg, H. A scale for the judgement of facial expressions. Journal of Experimental Psychology, 1941, 29, 497-510.

Shannon, C. & Weaver, W. The mathematical theory of communication. Urbana, Ill.: University of Illinois Press, 1949.

Steklis, H. & Harnad, S. From hand to mouth: some critical stages in the evolution of language. In origins and evolution of language and speech. Annals of the New York Academy of Sciences, 1976, 280, 445-455.

Stokoe, W. Sign language autonomy. In origins and evolution of language and speech. Annals of the New York Academy of Sciences, 1976, 280, 505-513.

Sumby, W. & Pollack, I. Visual contribution to speech intelligibility in noise. The Journal of Acoustical Society of America, 1954, 3, 212-217.

Searle, J. Speech acts. New York: Cambridge University Press, 1969.

Woodward, M. & Barber, C. Phoneme perception in lipreading. Journal of Speech and Hearing Research, 1960, 3, 212-223.

Zuckerman, M. Lipsets, M., Koivumaki, J. & Rosenthal, R. Encoding and decoding nonverbal cues of emotion. Journal of Personality and Social Psychology, 1975, 32, 1068-1076.