Video Transmission of American Sign Language and Finger Spelling: Present and Projected Bandwidth Requirements

GEORGE SPERLING

Abstract—**The bandwidth for communicating ASL sentences, isolated lists of ASL nouns, and finger-spelled names by a television type of raster scan was measured for a heterogeneous deaf population. Video sequences of a signer were shown to subjects at 60 frames/s (without interlace) with a raster composed of from 9 to 79 lines/frame corresponding to a bandwidth of from 1.1 to 86 kHz. Most subjects could interpret ASL sentences with little loss at a bandwidth of 21 kHz; the most expert ASL users received sentences at 40–50 percent correct at a bandwidth of 4.4 kHz. Even with the present nonoptimal raster, expert signers require a bandwidth for ASL communication that is just four times greater than the bandwidth required by experienced speakers for equivalent performance with auditory speech. With a more judicious choice of raster parameters, with better display conditions, and with better camera position, it seems quite probable that the present raster-bandwidth requirement could be lowered. Transmission at ordinary (3 kHz) telephone bandwidths will require more sophisticated picture coding, perhaps with different codes for ASL, finger spelling, and speechreading.**

INTRODUCTION

ABOUT two million Americans are severely deaf–to the point of being unable to understand speech even with a hearing aid [30]. Of these, about 200 000 were born deaf or became deaf before they learned a spoken language, about 410 000 became deaf before the age of 19 years, and most of the remainder became deaf in later life as an all-too-common concomitant of aging.

Throughout history, the deaf have evolved manual sign languages to communicate with each other. Perhaps because no widely accepted written form of any sign language has evolved,[1] these languages have not developed quite the range of vocabulary and sophistication of the most advanced spoken languages (see [32]). Nevertheless, the fact remains that American Sign Language (ASL, Ameslan)–the sign language now in common use in the U.S.A. and Canada–is perfectly adequate to enable deaf persons to communicate with each other on everyday, nontechnical subjects at about the same speed as hearing persons communicate in ordinary speech [2], [19]. ASL is in no sense a copy of English–it has its own distinctive grammar and modes of expression [3], [22], [34].

Words such as proper names or technical terms for which no sign yet exists in ASL can be expressed by means of finger spelling–a letter-by-letter rendition of the word by means of a stylized set of finger positions. The rate of finger spelling normally is several letters/s with skilled users approaching a rate of 10 letters/s. Nevertheless, finger spelling is much slower than either spoken language or ASL [4].

Sign languages enable the deaf to communicate among each other with great facility, in contrast to the difficulty with which the deaf communicate with the hearing community by means of reading lips and facial expressions, and by means of written messages. Perhaps for this reason, sign languages such as ASL have been perceived by the hearing community as contributing to the isolation of the deaf, and have frequently been ridiculed or even proscribed by educators of the deaf who favor an oral (speechreading) approach to communication. The author does not wish to enter here the oral-versus-manual controversy, but to note that–because it can be easily learned and greatly speeds communication–ASL is known to the majority of congenitally deaf adults regardless of their educational background [21, p. 187].

The aim of the present research is to solve the problem of putting the transmission facilities that already have been established for voice communication by telephone at the service of the deaf community. It is ironic that the telephone, which was invented by Alexander Graham Bell who also conducted extensive research on hearing aids for the deaf [11], has served almost exclusively the hearing community and has actually increased the isolation of the deaf community. For telecommunications involving the deaf, there are basically two kinds of devices under consideration; these are exemplified by teletypewriters [1] and video telephones (e.g., the American Picturephone and the English Viewphone) [6], [7], [13], [15], [25].

Teletypewriters are devices that enable a sender to transmit a typewritten message to a receiver who sees the characters displayed on a screen or produced on a teletypewriter. The teletypewriter is ideal for communication between deaf and hearing people–perhaps in conjunction with a voice channel for those deaf who retain intelligible speech. The main practical disadvantage of communication by teletypewriter is that typewriting is slow and effortful compared to voice or ASL communication. Nevertheless, in the absence of an alternative deaf/hearing telecommunication device, the development and deployment of reliable, low-cost teletypewriters is a critical necessity.

The video telephone station transmits a picture of the

Manuscript received December 30, 1980; revised October 27, 1981.

The author is with the Department of Psychology, New York University, New York, NY 10003, and Bell Laboratories, Murray Hill, NJ 07974.

[1] Notational systems for sign language have been proposed by Stokoe, Casterline, and Croneberg [35] and by Cohen, Namir, and Schlesinger [5], but these have not been widely adopted.

0090-6778/81/1200-1993$00.75

sender to the receiver by means of a television type of raster scan. The video telephone is an ideal medium for communication of ASL or finger-spelling [26]; the quality of the image in Picturephone and Viewphone is adequate even for speechreading. The main disadvantage of Picturephone and Viewphone is that they require a communication bandwidth of about 1 000 000 Hz—compared to about 3000 Hz for a telephone voice-communication channel [36]. (Ordinary American television requires 4 000 000 Hz.) The enormous bandwidth requirement for Picturephone and Viewphone—one video telephone communication channel could carry more than 300 voice telephone channels—not only makes current video telephones prohibitively expensive, it makes it impossible for these video telephones to use existing voice telephone transmission and switching facilities.

The present paper is concerned with the problem posed by Sperling [32]: what are the bandwidth requirements for ASL communication by video telephone, i.e., by a video telephone with lower bandwidth and consequently poorer picture quality than current video telephones? To what extent could such a video telephone use existing telephone channels to communicate ASL and finger spelling, and to what extent would new facilities be needed?[2]

PROCEDURE

Overview of the Procedure: To measure the bandwidth necessary for communication, we make a television recording of a signer producing sentences and other messages using ASL and finger spelling, and we test how accurately these messages can be transcribed by deaf subjects as a function of the bandwidth allocated for message transmission. This is a conservative method that overestimates bandwidth requirements because it is not interactive; the signer does not adjust his communication to compensate for a sign or word that the subject may have failed to understand, nor does the signer adjust his performance to emphasize or to compensate for aspects of ASL that may be poorly transmitted. On the other hand, it is an extremely convenient method because many persons can take exactly the same test and because scoring is relatively unambiguous. This objective determination (of the image's utility for communication) is much preferable to the subjective judgment of image quality that is traditionally used to evaluate processed images.

A reduced-bandwidth channel can be simulated by a reduced area of a full television picture. Television cassette recordings are produced that correspond to the different bandwidth channels; the appearance of ASL that has been degraded by passing through the various bandwidth channels can be demonstrated by simply inserting the appropriate cassette into a calibrated playback recorder and viewing the picture.

Bandwidth, Raster Scans: Television, Picturephone, and Viewphone use a raster scan. That is, a picture is represented as a series of lines called a raster. Each line is drawn from left to right so that the intensity at each point along the line represents the picture intensity at that point, and the lines are filled in from top to bottom to produce a complete picture. The American television picture-raster consists of 525 lines, and there are 30 entirely new pictures (called full frames) produced each second. (In fact, commercial television is slightly more complicated; a full frame is composed of two half-frames, the first of which contains only the odd-number raster lines (counting from the top) and the second of which contains the even-numbered raster lines (interlace). In American television there are actually 60 half-frames, equivalent to 30 full frames, per second.)

Bandwidth of a channel refers to the maximum number of cycles per second (hertz, Hz) that a channel can transmit. Other things equal, the amount of information a channel can transmit is proportional to its bandwidth. What makes the bandwidth of television so high is the large number of lines that must be represented each second (15 750) and the large number of discrete intensities that can be represented along each line.

Could a low-bandwidth picture with a raster composed of fewer, shorter lines, perhaps repeated less often each second, still serve adequately to communicate ASL? In principle, this question could be investigated by constructing video communication devices that operate on such rasters. In practice, this approach is expensive and technically difficult. There is an expedient alternative: to use a restricted, small area of a television screen (which obviously does contain fewer, shorter lines than the whole frame) to represent a reduced-bandwidth channel. By measuring the bandwidth of the whole television picture and measuring the fraction of the total area to which ASL transmission is restricted, one can determine precisely the bandwidth being allocated for ASL transmission.

The remainder of this Procedure section deals with the television recording procedure, the actual measurement of bandwidth (i.e., the calibration procedure), the construction of stimulus materials, the procurement of subjects, the vision-testing procedure, written instructions and questionnaires for the subjects, and the testing procedure.

Television Recording: Edgar Bloom, a deaf, retired chemist, served as signer. He stood behind a large black screen in which a 30.5 cm (12 in) wide by 45.7 cm (18 in) high rectangular aperture had been cut out to reveal his head and part of his shoulders (Fig. 1). Preliminary experiments without an aperture had shown that this size of aperture would be adequate. The signer succeeded—with one or two accidental exceptions—in compressing ASL signs that normally occupied an area extending to his waist, into the aperture area. He wore a dark shirt and jacket, and on his right hand, a dark cotton glove with white fingers, a glove that had been shown in the preliminary experiments to increase slightly the legibility of his signing at low bandwidths. Lights were arranged so that his hands appeared lighter than his face.

From the subject's point of view, the trial sequence was as follows. The signer was always visible on the television screen. At the beginning of a trial, a placard indicating the type of trial (sentence, noun, finger spelling) and the trial number was displayed on the screen for 5 s. The signer then rendered the stimulus item(s), after which he remained quiet for a suitable interval to allow subjects to transcribe their response. The transcription interval, chosen on the basis of preliminary experiments, was 15 s for noun triplets, 25 s for sentences, and 12 s for finger spelling.

[2] A preliminary account of these experiments is given by Sperling [33].

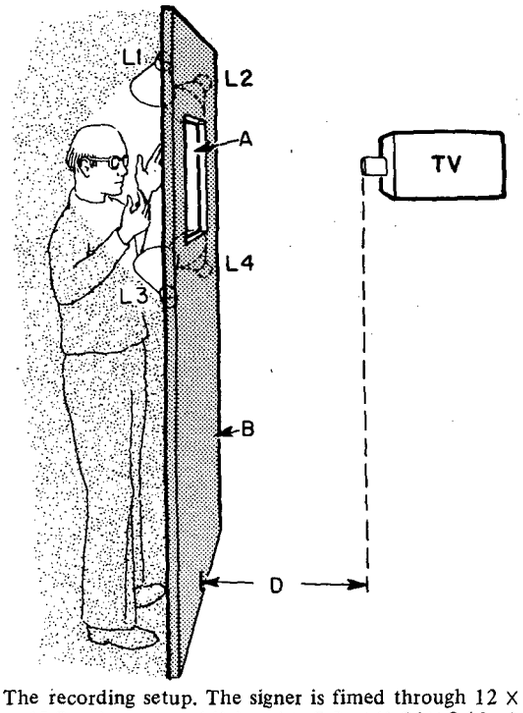


Fig. 1. The recording setup. The signer is fimed through 12 × 18 in (30.5 × 45.7 cm) aperture (*A*) in a 4 × 7 foot (1.24 × 2.13 m) screen with a television camera (*TV*) located at an aperture to lens distance (*D*). Lamps *L*1–*L*4 are arranged to emphasize the hands.

In the biggest picture (highest bandwidth) condition, the entire set of stimulus materials was filmed with an 8 mm lens at a lens-to-subject distance of 2 m. Successively smaller images (lower bandwidths) were produced as follows: 8 mm lens at 4 m; 4 mm lens at 4 m; 4 mm lens at 8 m. Lighting and other aspects of the recording remained the same, independent of the camera distance. A recording was made on a separate cassette for each bandwidth condition. The four conditions were then copied onto a single cassette that was used to run the subjects. (The extra cassette-copying operation was omitted in the calibration procedure; thus, the bandwidths given below are slight overestimates.)

Bandwidth Calibration: The television recording/playback system consisted of a Panasonic WV204P TV camera with an Apollo 4 mm (or 8 mm) lens, a Sony VO-2800 TV-cassette recorder (3/4 in), and a Conrac QQA 17/N monitor with a 17 in cathode ray tube. Calibration recordings were made of a test pattern—a 100 percent modulated, vertical square-wave grating at various distances from the camera (i.e., the test pattern consisted of alternating black and white bars). These recordings of the test pattern produced gratings with a different number of cycles/cm on the viewing screen, depending on the distance from camera to test pattern. The modulation depth (i.e., the contrast between dark and light bars) on the screen of the various size patterns was measured with a microphotometer. Coarse gratings were reproduced without any loss of modulation depth. A very sharp decrease in modulation depth occurred for gratings whose component lines on the screen were spaced at distances that corresponded to a transmission frequency of 2×10^6 Hz. The grating for which modulation depth dropped to 50 percent of the modulation of very coarse gratings—the half amplitude point—was 2.0 ± 0.1 MHz. Tests without the cassette recorder showed that the recorder itself was, by far, the limiting component; the remainder of the system easily passed 4 MHz. Measurements of various partially reflecting gray papers showed good linearity in gray-scale reproduction.

The television camera did not have the internal synchronization mechanism required to make interlaced half-frames. The net effect is simply equivalent to 60 frames/s; each frame has half the number of raster lines of American television (which has 30 frames/s). The absence of interlace in the display left dark spaces between the raster lines. These regularly alternating dark lines produce a horizontal grating with a high spatial frequency superimposed on the picture—an operation that is known to produce masking and to interfere with resolution of details [14], [24].

The actual sizes of the various picture conditions on the viewing screen and their nominal bandwidths are given in Table I. Photographs of the three highest bandwidth conditions are shown in Fig. 2. Because of the difficulties of static representation of dynamic television displays, and because of the problems inherent in photographic reproduction, the photographs in Fig. 2 appear to be of lower quality than do the corresponding dynamic displays.

Stimulus Materials: There were three kinds of stimulus materials: 1) short sentences rendered in idiomatic ASL, 2) common nouns (in sequences of three) rendered in ASL, and 3) last names of persons and city names rendered by finger spelling. The sentences and nouns were adapted from materials in a popular ASL textbook [23].

Sentences varied slightly in length from four to six signs. English transcriptions of some typical sentences were: "Please summarize last paragraph," "Rat runs across street," "Careless driving causes many accidents," "Modern art irritates many people," "You can't play football before Friday," etc.

Typical noun triplets were: "machine, battle, sweetheart," "science, brother, experience," "mirror, life, college," etc.

Last names for finger spelling had six or seven letters, and city names had mostly six to eight letters. The first set of ten items for finger spelling was: "Jensen," "Untamo," "Isacson," "Yoshida," "Preston," "Albany," "Burbank," "Madison," "Norfolk," "San Diego."

Four matched sets of stimulus items were constructed for use in the four bandwidth conditions to be tested. In order to avoid possible experimenter bias in assigning items to sets, all the stimulus items were ordered by length (within each category), and they were then assigned to the four sets, four at a time, by a random procedure.

A full set of stimuli for each bandwidth condition consisted of, in order, ten noun triplets, ten sentences (the first five were signed once and each sentence in the second five was repeated immediately after its first production), five finger-spelled person names, and lastly, five finger-spelled city names.

The precise ASL rendition of the nouns and sentences was worked out in advance by the signer, and one or two stimulus items that the signer felt he could not produce unambiguously were eliminated. The signer used "total communication," that is, he not only used ASL signs made with the hands but he enunciated each word to provide face movements that could be used for speech reading. Signing was somewhat slower than normal, with careful emphasis—in the style of a news-

TABLE I
PICTURE DIMENSIONS OF THE VARIOUS BANDWIDTH STIMULI

	Actual Picture Size Width × Height (cm)	Fraction of Full Picture Area	BW (Hz)	Number of Horizontal Raster Lines	Raster Spacing (cm)[b]
full picture	39.7 × 29.8[a]	1.0	2 000 000	262	
A	5.78 × 8.07	0.0357	86 300	79	0.58
B	2.86 × 4.28	0.00870	21 000	38	1.20
C	1.31 × 1.96	0.00182	4390	17	2.61
D	0.65 × 0.98	0.000455	1098	9	5.22

[a] Includes 12.6 percent of picture which is off the display screen.
[b] The distance between raster lines at the signer's aperture.

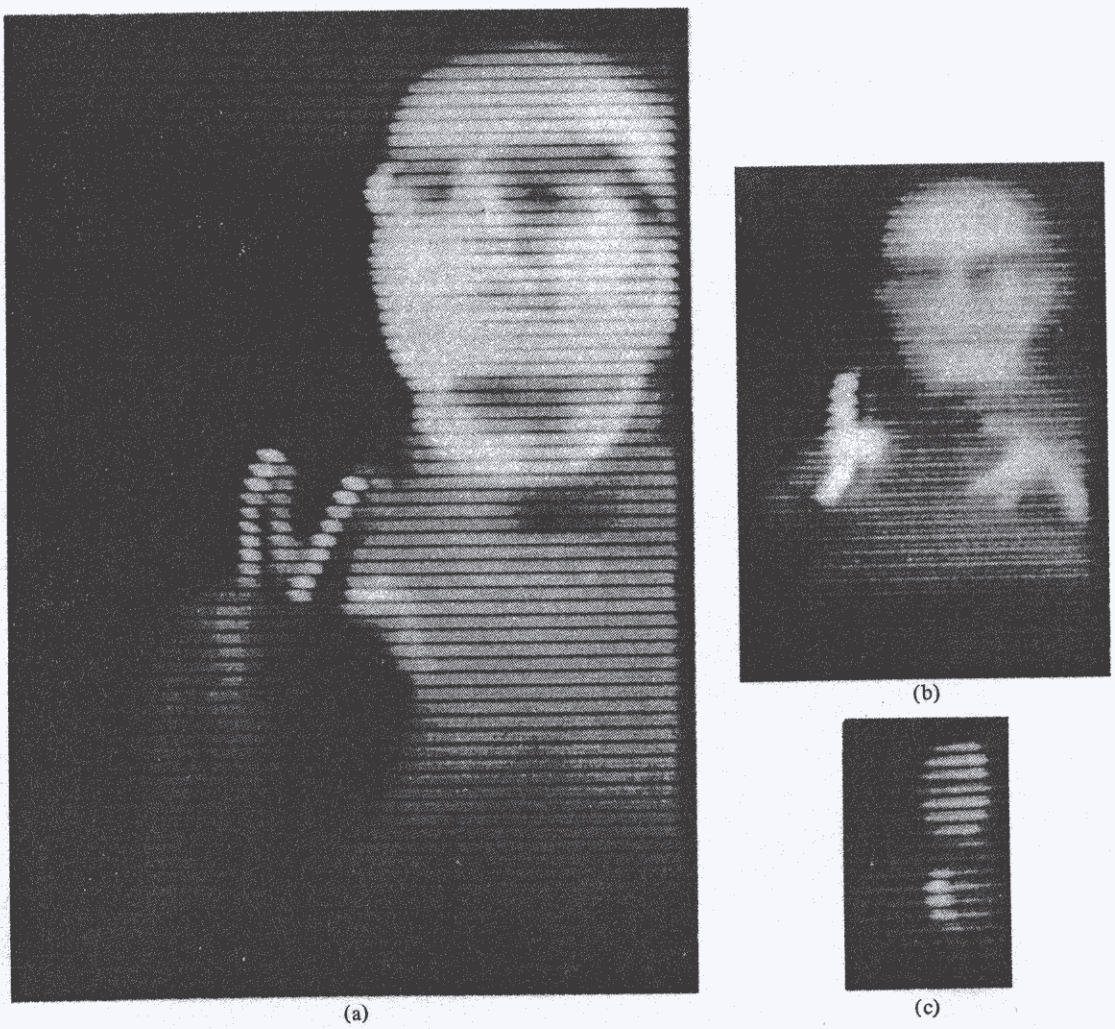


Fig. 2. Photographs of picture displays at three bandwidth conditions: (a) 86 300 Hz; (b) 21 000 Hz, (c) 4390 Hz. Exposure: 1/10 s, f/5.6, Polaroid ASA 3000 sheet film. The actual sizes of the three displays were 5.8 by 8.1 cm, 2.9 by 4.3 cm, and 1.31 by 1.96 cm. Because of quality losses in photography and in reproduction, the subjective quality of these photographs corresponds more nearly to the three lowest bandwidth conditions.

caster making a significant announcement. Sounds were not recorded. Noun triplets were signed at the rate of one noun/s. The average sentence rendition took about 5 s. Finger spelling was done at about 3–4 characters/s for the two highest bandwidth conditions and about 2–3 characters/s at the lower bandwidth conditions.

Subjects: In an attempt to obtain a representative sample of deaf persons, subjects were obtained through several sources. Two subjects were clerical employees at Bell Labs, two subjects were employees and one a graduate student at the NYU Deafness Research and Training Center, and 17 subjects were obtained through a call for subjects at a New York social club for deaf persons and through individual solicitations by several of the members of their friends. Because the tests were con-

ducted during normal working hours, this last group of subjects contained six housewives, five retired persons, a paraprofessional teacher, a clerk, a machine operator, a printer, an unemployed printer, and a disabled printer. It was not permitted to pay the Bell Labs subjects; the NYU subjects each received $4.00 for their participation. The remaining subjects received $7.00 as an incentive to participate and to compensate them for the time spent traveling.

Testing Procedures

Questionnaire: Subjects were asked to fill out a questionnaire that inquired about their age, occupation, education, severity of hearing loss, and usage of sign language.

Eye Test: Subjects were asked to copy one line of the smallest print they could read from an eye test chart (American Optical 1490) that contained samples of text ranging in size upward from 0.05 in (vertical height of capital letters). This was to ensure that their visual acuity did not limit their performance in looking at the smaller pictures.

Instructions: Subjects were seated at a table facing a television set at eye level. They could comfortably vary their viewing distance from about 30 cm (1 ft) to about 100 cm (3 ft). The room was dimly illuminated; there was a small directional lamp on the table to enable them to read instructions and write responses. They were given written instructions that outlined what they would see and instructed them to transcribe it in English—not necessarily good English—on a prepared answer sheet. They were specifically instructed to write down part of a sentence when they could not see it all, part of a finger-spelled word if they did not see all the letters, etc. With many subjects, it was necessary to repeat this instruction several times during the tests.

For most subjects, testing was terminated when they indicated—by writing nothing for several trials in a row—that they could not interpret the videotape.

Scoring: Responses were scored for the number of items correct. In the case of noun triplets, each noun counted as one item. Occasionally, half-credit was given as, for example, when a subject wrote "grandmother" instead of "mother." Sentences were scored as if they simply were composed of four, five, or six items. For finger spelling, the number of correctly written letters was counted, subject to the constraint that their order match the stimulus order. The score for each condition is given as a fraction of the total possible score.

There was very little guessing and relatively few incorrect responses. Subjects tended to leave blanks rather than guess. Even when subjects understood most of a sentence or most of the letters in a word, they often wrote nothing. It took repeated urging by the experimenter to persuade subjects to write the fraction of the message they had received. For example, in a finger spelling condition, one subject did not write even a single letter for six consecutive words and then wrote the next four words with only one minor spelling error. Thus, scores are underestimates of the utilizability of a channel.

Subsequently, it was discovered that two ASL signs ("telegram," "cards") had been made in a way that was unfamiliar to the subject population; these and a sentence partially out of the field of view were eliminated from the scoring. Two signs (in the biggest bandwidth condition) were ambiguous; subjects were given credit for either interpretation.

RESULTS AND DISCUSSION

The fraction of items correctly transcribed by each subject in each condition is shown in Fig. 3. The subjects are grouped into several graphs because of the difficulty of displaying data from more than about five subjects in one graph. The main overall result is obvious: as bandwidth decreases, accuracy decreases. This is true for all subjects and stimulus materials. Beyond the obvious result, matters are more complex.

Control Condition: For ASL signs, the 86 kHz picture can be regarded as completely intelligible, and a subject's score can be regarded as a close approximation to a noninteractive face-to-face conversation. Thus, a bandwidth of 86 kHz approximates infinite bandwidth for the present stimuli. Typically, when subjects could not transcribe a sign in the 86 kHz condition, they would imitate it perfectly and ask the experimenter what it meant.[3] By contrast, in low bandwidth conditions, when subjects failed to transcribe, they usually indicated that they were unable to *see* what was being signed—it was "too small" or "too blurred." The 86 kHz condition thus serves as a *control condition* for the other bandwidth conditions. Errors were due primarily to unfamiliarity with the particular ASL signs used, memory lapses (in the transcription of three nouns or a four to six item sentence) or lapses of attention—subjects were looking away when the stimulus occurred. These kinds of errors are more or less equally likely to occur in all conditions—higher as well as lower bandwidths—and therefore 86 kHz is a suitable reference condition.

In finger spelling, errors in the 86 kHz condition occurred primarily because the signer signed "too fast" (about 3-4 characters/s). This was an invariable complaint to the experimenter by subjects. Had the finger spelling been slower, there is little doubt that even these subjects could have transcribed it correctly, and future tests will contain finger spelling exercises produced at fast and at slow rates. But, since finger spelling usually is moderately fast, it is legitimate to ask whether performance of the weaker subjects in the highest bandwidth might have been improved by still more bandwidth. This question cannot be answered on the basis of the present data. The author's guess is that bandwidth beyond 86 kHz would have improved finger spelling performance slightly for the weaker subjects. Fast finger spelling was not tested because many subjects could not have transcribed it; on the other hand, fast finger spelling undoubtedly would benefit from more bandwidth.

Performance at Reduced Bandwidths: Fig. 4 shows the performance of each subject relative to his/her performance in the 86 kHz control condition. (The two subjects who were not proficient in ASL are omitted from these and subsequent graphs.) The median performances of the top quartile, middle 50 percent, and lowest quartile of the subjects are shown. Most subjects transcribe ASL sentences and nouns at 21 kHz with about 90 percent ± 10 percent of their accuracy in the

[3] The ability to "shadow" a signer may be a better test of intelligibility than the ability to transcribe a message. But it is inconvenient to administer and to score a shadowing test.

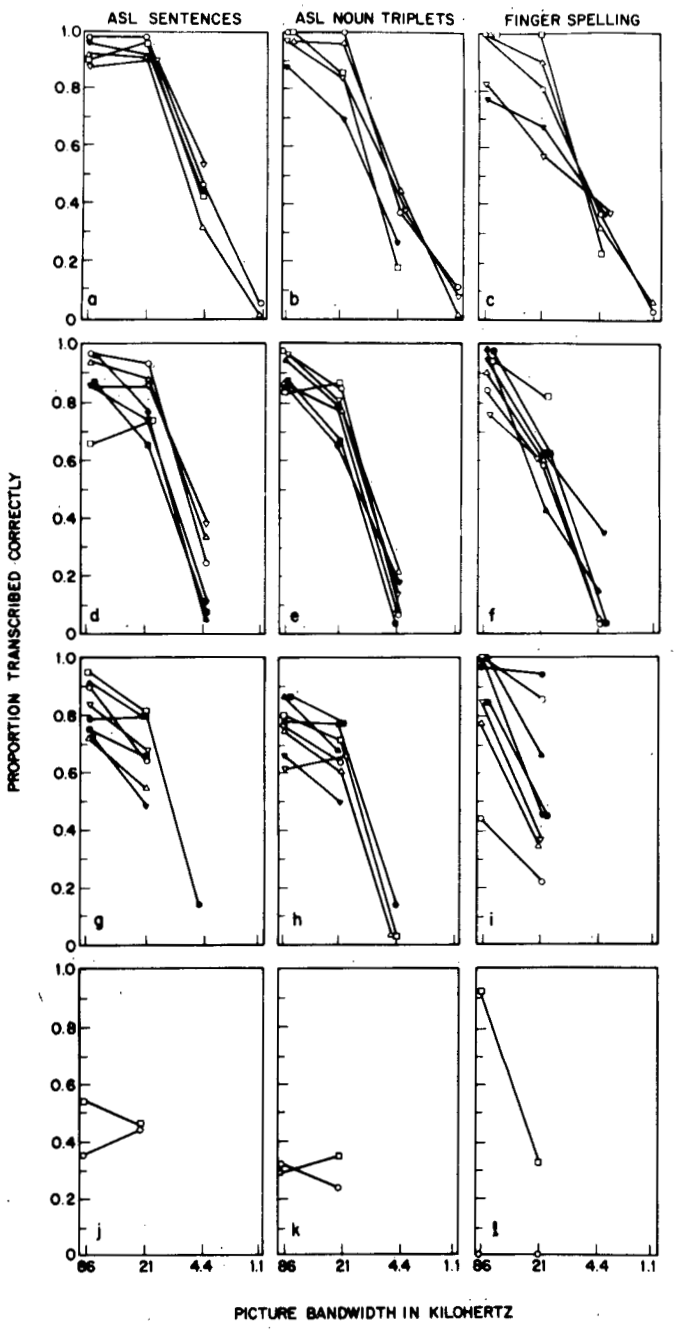


Fig. 3. Absolute proportions of correctly transcribed items as a function of picture bandwidth for American Sign Language sentences, noun triplets, and for finger-spelled names. Each point represents the score for one subject; where no point is indicated, the subject did not complete the task. Scores for each individual subject in each task are connected by lines. Each subject is represented by similarly coded points in one row of graphs. The most accurate ASL-transcribing subjects are represented in the top row of graphs; the two subjects who had a very incomplete knowledge of ASL are represented on the bottom row.

control condition. Below 21 kHz, performance drops precipitously, although the top quartile maintain around 40-45 percent of control accuracy out to 4.4 kHz. One third of the subjects are unable to transcribe anything from stimuli at 4.4 kHz bandwidth or less.

With finger spelling, the results for the top quartile are much the same. The top quartile maintains 90 percent of control accuracy at 21 kHz and about 40 percent at 4.4 kHz. The median subject suffers more from bandwidth reduction with finger spelling than ASL, and performance is down to 70 percent of control at 21 kHz. The lowest quartile is even more affected, slipping to about 45 percent of control performance at 21 kHz. From the parallelism and separation of the curves of Fig. 4, we observe that the top quartile subjects can maintain their performance to about one half the bandwidth of the bottom quartile subjects.

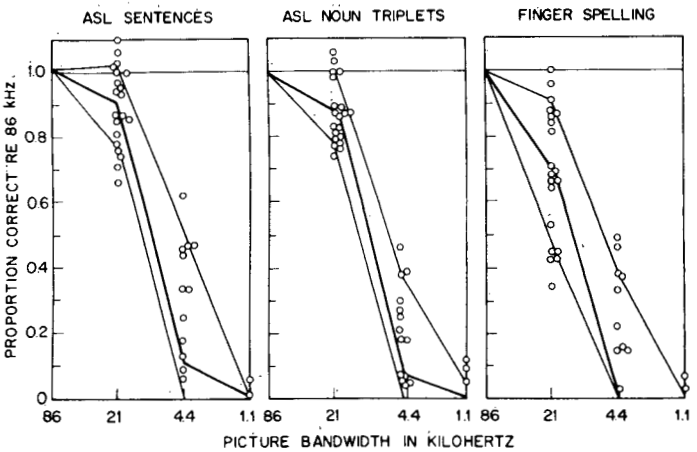


Fig. 4. Relative proportions: the proportions of correctly transcribed items as a function of picture bandwidth *relative* to the control condition (86 kHz) for ASL sentences, noun triplets, and finger-spelled names. Each point represents one subject's performance in one task condition. Points that would be superimposed are spaced apart horizontally to facilitate the display. The middle line in each graph indicates the median for all subjects; the upper line represents the median of the top quartile (87.5 percent), the lower line represents the median of the lowest quartile (12.5 percent); i.e., 75 percent of subjects fall between the upper and lower lines.

Comparison of the Different Stimulus Materials: Fig. 5 shows the average performance for nouns, sentences signed once, sentences signed twice, finger-spelled family names, and finger-spelled city names. Sentences signed twice are reported about 3 percent more accurately than sentences signed once—a very slight improvement. This indicates that simply repeating a misunderstood ASL sentence in a low bandwidth channel would probably not enable a receiver to better understand it (i.e., the "noise" is correlated). Overcoming low bandwidth requires paraphrasing, slower signing, emphasis, or some other alteration of the message.

Finger-spelled city names are transcribed with 15-20 percent greater accuracy than family names. This probably reflects a familiarity/redundancy factor in transcribing finger spelling—a classical word-frequency effect [16]. Presumably, even though the subjects had difficulty in reporting partial information—component letters of names not fully understood—they occasionally were able to use this partial information to generate intelligent guesses [28], [29]. There is no evidence whatever that the component letters were produced any differently by the signer in family than in city names.

Vision Test: Four of 22 subjects were unable to read the smallest type on the vision test but were able to read larger type; these four ranged in age from 49 to 68. Two of these subjects also happened to be the two who were least proficient

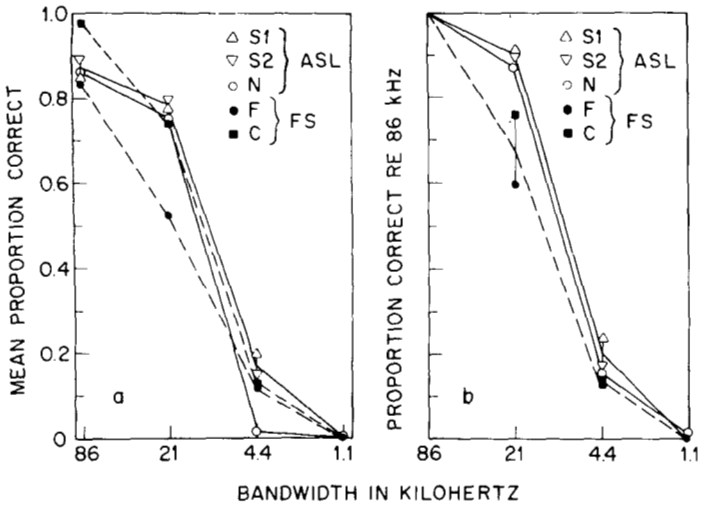


Fig. 5. Mean proportions of correctly transcribed items shown separately for all materials and conditions tested as a function of picture bandwidth. Means are taken over the 20 subjects who knew both American Sign Language and finger spelling. The ordinate in (a) is the (absolute) proportion correct; in (b) it is the proportion correct *relative* to 86 kHz control condition. *S*1 = ASL sentences signed once; *S*2 = ASL sentences signed twice; *N* = ASL noun triplets; *F* = finger-spelled family names; *C* = finger-spelled city names. Lines are drawn through the overall sentences means and [except in (a)] through the overall finger spelling means.

in ASL, including the only subject in the study who could not read finger spelling. These data are shown in Fig. 3(j), (k), and (l). It is interesting to note in Fig. 3 that the slightly below average vision of these two subjects did not impair their ability to interpret ASL in the smaller pictures (relative to other subjects) since these two subjects showed no decrement in ASL performance in the 21 kHz condition relative to the 86 kHz condition. Except for their lower accuracy, the ASL data of these two subjects are not different in any obvious way from the data of subjects with better vision. There is an unusually large decline in finger spelling performance at 21 kHz for the subject who was able to read finger spelling. This is probably not simply due to impaired vision because the two proficient vision-impaired subjects showed perfectly normal declines in finger spelling and normal ASL data. In all, we must conclude that slight visual impairments do not impair the ability of subjects to interpret these pictures. We account for this simply by assuming that all the low bandwidth pictures are much worse—i.e., even more blurred—than the subjects' vision.

The Enormous Range of ASL Proficiency: Compared to spoken English, the range of individual variation in the ability to communicate in ASL is larger, i.e., in terms of vocabulary size and rate of communication. Furthermore, there seems to be a larger range than in spoken English of regional dialect differences, so that some signs commonly in use in New York might be misunderstood in New Jersey. In part, such differences may arise because there is no widely accepted written form of sign language; in part, because most of the subjects in this experiment were trained in the oral tradition, and they never received any formal training in ASL.

The range of individual variation in finger-spelling proficiency also is large, perhaps because of differences in the amount of formal training. Schools in the oral tradition rely heavily on finger spelling to supplement lip reading.

The differences in individual proficiency are not fully expressed in the control condition (86 kHz) scores because of a ceiling effect. The materials were designed to be widely understood under good viewing conditions and this was the case. The full range of individual ASL proficiencies becomes manifest when the picture is degraded. For example, with sentences at 21 kHz, the best ASL subjects show no decrement from their near perfect performance at 86 kHz and they still transcribe about 40 percent correctly even at 4.4 kHz. Most other subjects show some decrement at 21 kHz; about one third of the subjects give up completely and do not transcribe anything at 4.4 kHz.

The most proficient subjects in ASL are not always the most proficient subjects in finger spelling. The correlation across subjects between ASL and finger spelling scores is $r = 0.615$ ($\pm 1\sigma = -0.17, +0.12$) for 21 kHz and $r = 0.113$ (± 0.23) for 86 kHz. The reduced correlation for 86 kHz is due to the ceiling on scores. Differences in educational background would tend to negatively correlate ASL and finger spelling proficiency (different schools emphasize different skills). General language ability and level of education would positively correlate ASL and finger spelling proficiency.

In spoken language, the relation of language skill to the ability to communicate over noisy channels becomes manifest when people attempt to use second languages on telephones or in listening to noisy radio broadcasts. Language deficits that do not impair face-to-face communication can become incapacitating under these less than optimal conditions. For the deaf subjects, the ability to interpret ASL and finger spelling under low bandwidth conditions is itself a quick test of proficiency in these language forms.

Implications for Low-Bandwidth Video Telecommunication

Comparison of Bandwidth Requirements for Visual and Auditory Communication: It is interesting to compare the 4.4 kHz bandwidth required for 40-50 percent intelligibility of visual ASL communication with the bandwidth of about 1.4 kHz that is needed to achieve a comparable intelligibility level with auditory speech communication [12]. Thus, even with the present nonoptimal raster parameters, expert signers require just four times more bandwidth for raster-encoded visual ASL communication than experienced speakers do for pressure-waveform encoded auditory speech communication. The performance of the deaf subjects as a whole at 21 kHz with the present raster is roughly equivalent to the performance of normal hearing subjects with a 3 kHz, ordinary telephone line.

The Present Raster Parameters Are Nonoptimal: The present method of simulating low-bandwidth channels by using a restricted area of a television picture gives an overestimate of the required bandwidth because it does not optimize the parameters of the low-bandwidth channel. Except for the absence of interlace, the parameters are those of American television; thus, the picture code was a raster-scan with 60 frames/s. The extent to which intelligibility could be improved by changing these parameters is a question for further research.

For example, could a slower frame rate with a finer raster (or with a double or triple interlaced raster) give better results? Certainly, because the loss of resolution of finger position was the main problem at low bandwidths. Is there a choice of a vertical or horizontal scan direction and of the number and the spacing of scan lines that would improve intelligibility? Probably. These questions themselves assume a uniform raster scan. Perhaps there ought to be a higher density of raster lines in some parts of the picture.

To return to the practical problem of ASL transmission: the simple expedient of moving the camera to within about 75 cm (30 in) of the signer (so that when he makes signs in front of the body, the fingers are closer to the lens and hence produce larger images) would do much to improve finger resolution—the main limiting factor at low bandwidths with a raster scan.

Specialized Picture Codes: Because the required bandwidth is so low, somewhat more complex picture codes than raster scans could be implemented in real time. The advent of low-cost microprocessors makes these alternatives economically feasible.

A specialized picture code for ASL would incorporate constraints such as that the head and shoulders move relatively slowly and predictably and do not need to be resolved very accurately. The hands and fingers move quickly and require fine resolution—although it usually is only one primary hand that requires resolution.

Finger spelling requires resolution just of the fingers of one hand. Because finger spelling proceeds much faster than ASL, a picture code specialized for finger spelling might use more frames/s than a code for ASL.

Speechreading ("lipreading") requires resolution of the mouth, primarily, and of the lower part of the face [17] including the cheeks [31]. The ability to resolve the shape of the lips and to see the teeth and the tongue through the lips is important to speechreaders. Thus, a specialized picture code for speechreading requires better contrast resolution but less spatial resolution than do codes for ASL and finger spelling. It has been shown experimentally [9] that a degraded auditory voice channel and a degraded visual channel (of the face)—each of which individually are insufficient—together can produce intelligible speech communication. Since speechreading is particularly helpful to persons who retain some residual hearing and who have good speech (predominantly older persons), it would be desirable to combine acoustic channels for voice communication with visual channels for speechreading to meet the needs of this hearing-impaired group.

For communication between congenitally deaf and hearing persons, as well as among the deaf, a serial, character transmitting device, such as a teletypewriter (TTY), has proved extremely useful (see the Introduction). When a TTY is used for dynamic interpersonal interaction, the character rate is so low that TTY transmission could be made available as a background mode added onto any of the video transmission schemes considered above. For example, persons communicating by TTY could simultaneously have available dynamic head-and-shoulders views of each other. Alternatively, a pure TTY system could be designed to utilize available bandwidth effectively, for example, by using a local memory to accumulate longer character strings and thereby confining utilization of the transmission channel to brief high-speed bursts of accumulated strings.

The optimal degree of specialization of specialized picture codes is itself something that still needs to be worked out. Should finger spelling, ASL, and speechreading information be transmitted by the same or by different codes? If by different codes, how different? At one extreme of specialized codes, all that varies from code to code is parameters. For example, raster scan codes can trade spatial resolution for temporal resolution simply by trading off the number of raster lines in a frame against the number of new frames/s. Though somewhat specialized, such codes are general-purpose in the sense that they can transmit a wide range of pictorial material.

At the other extreme of specialization, there are codes that incorporate almost completely the constraints of the material being transmitted. For example, Erber demonstrated useful communication of information essential to speechreading by reconstructing outline lips based on a few parameters extracted from an actual speaker's lips [8] or from the acoustical voice signal [10]. Similarly, Montgomery's [20] subjects were able to lipread computer-reconstructed outlines of an actual speaker's lips. An analogous code for ASL or finger spelling would involve reconstructing an image of a hand based on transmitted information about only a few points on the hand. An interesting variant of this procedure based on Johannson's paradigm [18] was recently carried out by Poizner, Bellugi, and Lutes-Driscoll [27] and by Tartter and Knowlton [36]. All the viewer sees is a dozen spots on the fingers and wrist of each hand; the reconstruction of the actual hand and finger movements takes place in the viewer's mind.

Multiple Specialized Codes: Obviously, the development of specialized picture codes for speechreading and for manual sign languages is a fascinating problem for research. With the advent of low-cost microcomputers, video communication need not be restricted to just one picture code or even one class of codes. It would be quite reasonable for each of the connected parties to have several picture coding algorithms available, and to use the initial transmission to agree upon the codes to use for subsequent communication. Such modes of video communication might have applicability beyond the hearing-impaired, especially if an appropriate set of agreed-upon coding algorithms were available. As a trivial example, text or pictorial matter could be sent at high resolution at a very slow frame rate, dynamic head-and-shoulder portraits at medium resolution with medium frame rates, and finger spelling at high frame rates.

Relation to the Telephone: The present upper-bound estimate of required bandwidth (21 kHz) is too high for telephone circuits. But with minor improvements—a more judicious choice of raster parameters (interlace, slower frame rate), better display conditions (no blank space between raster lines in display, larger display-picture size), and better camera position—it seems quite probable that image quality comparable

to the 21 kHz raster that served ASL quite satisfactorily in the present tests could be achieved at lower bandwidths. The most facile subjects, those who achieved 40-50 percent intelligibility scores at 4.4 kHz under the present conditions, could almost certainly use a 10 kHz video channel for satisfactory communication even with present nonoptimal raster-parameters. However, to permit generally useful ASL communications at telephone bandwidths of 3 kHz requires approximately a sevenfold compression of information, and that degree of compression requires major improvements in image coding.

In the Future: Beyond fulfilling an immediate telecommunication need for the deaf and the partially deaf, video communication would also provide an impetus for acquiring greater skills in ASL and in three visual modes of English: finger spelling, speechreading, and teletypewritten communication. Video communication among the congenitally deaf would provide a strong impetus for development and standardization of the ASL language. These incentives for the further development of visual languages on one hand, and for the increased acquisition of language skills by the users on the other, might in the long term become the most significant benefit of video telecommunication for the deaf.

SUMMARY AND CONCLUSIONS

A heterogeneous group of congenitally deaf subjects was tested for the ability to transcribe American Sign Language and finger spelling from reduced television displays at bandwidths of 86 kHz, 21 kHz, 4.4 kHz, and 1.1 kHz.

For the median subject, ASL sentence intelligibility drops to 90 percent at 21 kHz and to 10 percent at 4.4 kHz.

The top quartile of subjects shows 100 percent ASL intelligibility at 21 kHz and about 45 percent intelligibility at 4.4 kHz.

Finger spelling is more sensitive than ASL to bandwidth reduction; median intelligibility drops to 70 percent at 21 kHz.

Modest vision impairments do not affect performance.

Repeating sentences twice produces only a negligible increase in intelligibility compared with sentences produced once.

With the present choice of television parameters for the raster scan, 21 kHz is satisfactory for communication by ASL and it is marginally adequate for finger spelling.

Minor improvements, such as shortening the signer-to-camera distance, optimizing the raster parameters, and improving the display conditions, would probably enable the same performance as 21 kHz at somewhat lower bandwidths.

Transmission at telephone bandwidths will require more sophisticated picture coding, perhaps with different codes for ASL, finger spelling, and speech reading.

ACKNOWLEDGMENT

Thanks to E. Bloom for freely donating many hours of his spare time to serve as signer; to K. Knowlton for his help in the selection of stimulus items; to P. Bricker for the loan of the television camera and recorder; and to Prof. J. Schein, Director NYU Deafness Research and Training Center, for arranging the use of the center's equipment and facilities for preliminary experiments, and for assistance in finding subjects.

REFERENCES

[1] P. A. Bellefleur, "TTY communication: Its history and future," *The Volta Review*, vol. 78, pp. 107–112, 1976.

[2] U. Bellugi and S. Fischer, "A comparison of sign language and spoken language," *Cognition*, vol. 1, pp. 173–200, 1972.

[3] U. Bellugi and E. S. Klima, "Structural properties of American Sign Language," in *Deaf Children: Developmental Perspectives*, L. Liben, Ed. New York: Academic, 1978, pp. 43–68.

[4] H. Bornstein, "Sign language in the education of the deaf," in *Sign Language of the Deaf*, I. M. Schlesinger and L. Namir, Eds. New York: Academic, 1978, pp. 333–361.

[5] E. L. Cohen, L. Namir, and I. M. Schlesinger, *A New Dictionary of Sign Language*. The Hague, The Netherlands: Mouton, 1977.

[6] T. V. Crater, "The Picturephone system: Service standards," *Bell Syst. Tech. J.*, vol. 50, pp. 235–269, 1971.

[7] I. Dorros, "The Picturephone system: The network," *Bell Syst. Tech. J.*, vol. 50, pp. 221–233, 1971.

[8] N. P. Erber and C. L. De Filippo, "Voice/mouth synthesis and tactual visual perception of pa ba ma," *J. Acoust. Soc. Amer.*, vol. 64, pp. 1015–1019, 1978.

[9] N. P. Erber, "Auditory-visual perception of speech with reduced optical clarity," *J. Speech and Hearing Res.*, vol. 22, pp. 212–223, 1979.

[10] ——, "Real-time synthesis of optical lip shapes from vowel sounds," *J. Acoust. Soc. Amer.*, vol. 66, pp. 1542–1544, 1979.

[11] M. D. Fagan, Ed., *A History of Engineering and Science in the Bell System: The Early Years (1825-1925)*, Bell Laboratories, Murray Hill, NJ, 1975.

[12] N. R. French and J. C. Steinberg, "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Amer.*, vol. 19, pp. 90–119, 1947.

[13] G. A. Gerrard and J. E. Thompson, *Radio Electron. Eng.*, vol. 43, p. 201, 1973.

[14] N. Graham, "Spatial frequency channels in human vision: Detecting edges without edge detectors," in *Visual Coding and Adaptability*, C. S. Harris, Ed. Hillsdale, NJ: Erlbaum, 1980, pp. 215–262.

[15] C. F. J. Hillen, "The face to face telephone," *Post Office Telecommun. J.*, vol. 24, pp. 4–7, 1972.

[16] D. Howes, "On the relation between intelligibility and the frequency of occurrence of English words," *J. Acoust. Soc. Amer.*, vol. 29, pp. 296–305, 1957.

[17] J. Jeffers and M. Barley, *Speechreading (Lipreading)*. Springfield, IL: Charles C Thomas, 1971.

[18] G. Johansson, "Visual analysis of biological motion and a model for its analysis," *Perception and Psychophysics*, vol. 14, pp. 201–211, 1973.

[19] E. Klima and U. Bellugi, *The Signs of Language*. Cambridge, MA: Harvard Univ. Press, 1979.

[20] A. A. Montgomery, "Effects of consonantal context on vowel lipreading," *J. Acoust. Soc. Amer.*, vol. 69, p. S122, 1981.

[21] D. F. Moores, "Current research and theory with the deaf: Educational implications," in *Deaf Children: Developmental Perspectives*, L. Liben, Ed. New York: Academic, 1978, pp. 173–193.

[22] L. Namir and I. M. Schlesinger, "The grammar of sign language," in *Sign Language of the Deaf*, I. M. Schlesinger and L. Namir, Eds. New York: Academic, 1978, pp. 97–140.

[23] T. J. O'Rourke, *A Basic Course in Manual Communication*, 2nd ed. Silver Spring, MD: Nat. Assoc. of the Deaf, 1973.

[24] A. Pantle and R. Sekuler, "Size detecting mechanisms in human vision," *Science*, vol. 162, pp. 1146–1148, 1968.

[25] D. E. Pearson, *Transmission and Display of Pictorial Information*. New York: Wiley, 1975.

[26] ——, "An experimental visual telephone for the deaf," *Television—J. Roy. TV Soc.*, vol. 16, pp. 6–10, 1978.

[27] H. Poizner, U. Bellugi, and V. Lutes-Driscoll, "Perception of

American Sign Language in dynamic point-light displays," *J. Exp. Psychol.: Human Perception and Performance*, vol. 7, pp. 430–440, 1981.

[28] I. Pollack, "Message uncertainty and message reception," *J. Acoust. Soc. Amer.*, vol. 31, pp. 1500–1508, 1959.

[29] H. B. Savin, "Word-frequency effect and errors in the perception of speech," *J. Acoust. Soc. Amer.*, vol. 35, pp. 200–206, 1963.

[30] J. D. Schein and M. T. Delk, Jr., *The Deaf Population of the United States*. Silver Spring, MD: Nat. Assoc. of the Deaf, 1974.

[31] J. S. Scheinberg, "Analysis of speechreading cues using an interleaved technique," *J. Commun. Disorders*, vol. 13, pp. 489–492, 1980.

[32] G. Sperling, "Future prospects in language and communication for the congenitally deaf," in *Deaf Children: Developmental Perspectives*, L. Liben, Ed. New York: Academic, 1978, pp. 103–114.

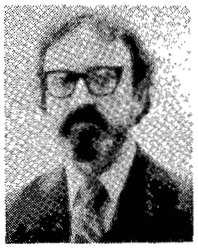
[33] ——, "Bandwidth requirements for video transmission of American Sign Language and finger spelling," *Science*, vol. 210, pp. 797–799, 1980.

[34] W. C. Stokoe, *Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf* (Studies in Linguistics, Occasional Papers: 8), G. L. Trager, Ed. Buffalo, NY: Univ. Buffalo Press, 1960.

[35] W. C. Stokoe, D. C. Casterline, and C. G. Croneberg, *A Dictionary of American Sign Language on Linguistic Principles*. Washington, DC: Gallaudet College Press, 1965.

[36] V. C. Tartter and K. C. Knowlton, "Perceiving sign language from an array of 27 moving spots," *Nature*, vol. 289, pp. 676–678, 1981.

[37] Technical Staff, Bell Labs., *Transmission Systems for Communications*, Rev. Ed. 4, Western Electric Co., Winston-Salem, NC, Tech. Pub., 1971.



George Sperling received the B.S. degree in mathematics from the University of Michigan, Ann Arbor, in 1955, the M.A. degree from Columbia University, New York, NY, in 1956, and the Ph.D. degree from Harvard University, Cambridge, MA, in 1959, both in experimental psychology.

He is Director of the Human Information Processing Laboratory at New York University, New York, where since 1970 he has been Professor of Psychology. Since 1959, he has also been a member of the Technical Staff of the Acoustical and Behavioral Research Center at Bell Laboratories, Murray Hill, NJ. Along with the above positions, he has been a Visiting Professor at New York University (1962–1963), Duke University, Durham, NC (1964), Columbia University (1964–1965), the University of California at Los Angeles (1967–1968), University College, London, England (1969–1970), the University of Western Australia, Nedlands (1972), and the University of Washington, Seattle (1977). He is the author of over 50 papers that deal mostly with mathematical and experimental analyses of human perception and information processing.

Prof. Sperling was awarded the Gomberg Chemistry Prize and was elected to Phi Beta Kappa, Phi Eta Sigma, Phi Kappa Phi, and Sigma Xi while at the University of Michigan. He has been elected a Fellow of the American Psychological Association, the Optical Society of America, and the American Association for the Advancement of Science. He has been elected to the Society of Experimental Psychologists and to the Executive Boards of the Eastern Psychological Association, the National Conference for On-Line Computing in Psychology, and the Society for Mathematical Psychology. He is a member of the Psychonomic Society and Founder of the Annual Interdisciplinary Conference (since 1975). In 1969–1970 he was awarded a Guggenheim Fellowship.