Bandwidth Requirements for Video Transmission of American Sign Language and Finger Spelling

Abstract. *Although current video communication schemes use a bandwidth on the order of 1 megahertz, the bandwidth required for video communication of American Sign Language by a simple raster scan is only approximately 20 kilohertz.*

Alexander Graham Bell invented the telephone in the course of his research on a hearing aid for the deaf (*1*), but the telephone has primarily served the hearing community. With more than 2 million severely deaf Americans who are unable to understand speech even with a hearing aid (*2*), it is appropriate for current research to find a means of adapting the transmission facilities already established for voice communication by telephone to the needs and habits of the deaf community. The most common methods of linguistic interaction among the congenitally deaf are American Sign Language (ASL), finger spelling (used where no sign exists in ASL, such as for proper names or technical terms), and speech reading (*3*).

Two existing devices that provide telecommunication for the deaf are the teletypewriter and the video telephone. Teletypewriters (*4*) enable a sender to transmit a typewriter message to a receiver, who sees the characters displayed on a screen or produced on another teletypewriter. The teletypewriter is useful for communication between deaf and hearing people, perhaps in conjunction with a voice channel for those deaf who retain intelligible speech. But the teletypewriter has a practical disadvantage: communication is slow and effortful compared with voice or ASL communication, which is about as fast as voice (*5*). The video telephone is far more attractive than the teletypewriter to many deaf persons for communication among themselves (*6*).

The American video telephone [Picturephone (*7*)] and the British version [Viewphone (*8*)] both transmit a picture of the sender to the reader by means of a television raster scan. Unfortunately, Picturephone and Viewphone require a communication bandwidth of 10^6 Hz [compared with 3000 Hz (*9*) for a telephone voice communication channel]. Their enormous bandwidth appetite not only makes them unsuitable for existing telephone transmission and switching facilities, but it makes the development of video telephone facilities economically unattractive. Present research is aimed at answering the question posed by Sperling (*10*): What are the minimum requirements for a video telephone having a lower picture quality than Picturephone or Viewphone but, consequently, a lower bandwidth so that it could utilize existing telephone channels?

Even if the technical resources were available, it would not be necessary to actually build a low-bandwidth transmission channel in order to discover its properties; it can be simulated by using ordinary television. American television uses a raster scan of 525 lines per picture, produces 30 full pictures per second, and has a bandwidth of 4×10^6 Hz (*11*). A small, rectangular subarea of the full picture represents a smaller channel; the subarea uses a fraction of the full bandwidth corresponding to the fraction of the full area it actually occupies. By recording a signer who is communicating in ASL or finger spelling and by confining the recording to a known, small area of a television picture, the bandwidth allocated to the ASL transmission can be readily computed. Of course, the full television transmission and recording system must first be calibrated to determine its true bandwidth (*12*).

The signer for this study stood behind a screen with a 12 by 18 inch (30.5 by 45.7 cm) aperture and produced ASL and finger spelling within the area of the aperture. Television recordings of the signer were made at four combinations of lens and camera distances to produce four bandwidth conditions: 86,000, 21,000, 4,400, and 1,100 Hz (Fig. 1) (*13*).

Four equivalent sets of stimulus materials based on vocabulary in a popular ASL textbook (*14*) were recorded. Each set contained (i) 30 nouns (for example, "son, coffee, morning"; "pants, people, face"; "autumn, paper, parents") and ten sentences (such as "Your book impresses me." "I need another stamp." "Sleepy girl doesn't want bath tonight.") produced in ASL and (ii) five finger-spelled family names and five finger-spelled city names (such as "Gilstrap," "Quigley," "El Paso," "Elmhurst").

In order to secure a heterogeneous sample of deaf subjects, tests were conducted at three places: two clerical employees were tested at an industrial laboratory; two employees and a graduate student were tested at the New York University Deafness Research and Training Center; and 17 subjects who responded to a call for subjects at a Brooklyn, New York, social club were tested at another New York University location. This last group of subjects consisted primarily of older persons and included six housewives, five retired persons, a teacher's assistant, a clerk, a machine operator, a printer, an unemployed printer, and a disabled printer. One housewife and one retired man did not know ASL and finger spelling; their data are omitted from the analysis.

Subjects were tested individually. They viewed the television display at a distance of from 1 to 3 feet (as they preferred), and after each triplet of nouns, each sentence, or each name, wrote down what they perceived of the transmission. The written responses were scored for the fraction of items reported correctly.

As bandwidth decreased, intelligibility decreased (Fig. 2). The highest bandwidth (86 kHz) was the control: errors that occurred at this bandwidth were attributable primarily to unfamiliarity with the particular ASL sign used, to lapses of attention and memory, and, in the case of finger spelling, to the inability of some

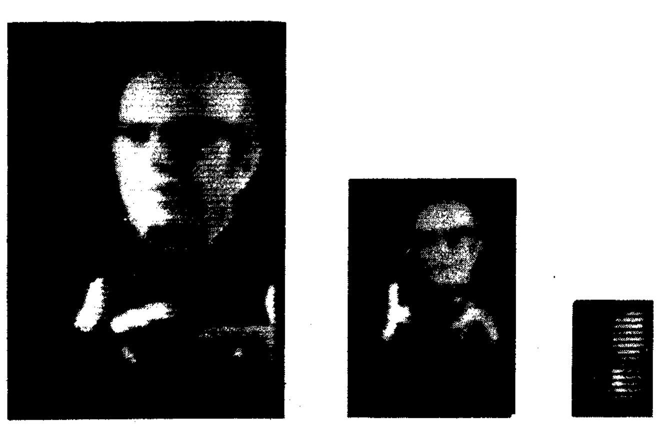

Fig. 1. Photographs of the television display at the three highest bandwidth conditions. (A) 86,000 Hz. (B) 21,000 Hz. (C) 4,400 Hz. Exposure: 1/10 second. The actual sizes of the three displays were 5.8 by 8.1 cm, 2.9 by 4.3 cm, and 1.31 by 1.96 cm. Because of quality losses in photography and in reproduction, the subjective quality of these photographs corresponds more nearly to the three lowest bandwidth conditions.

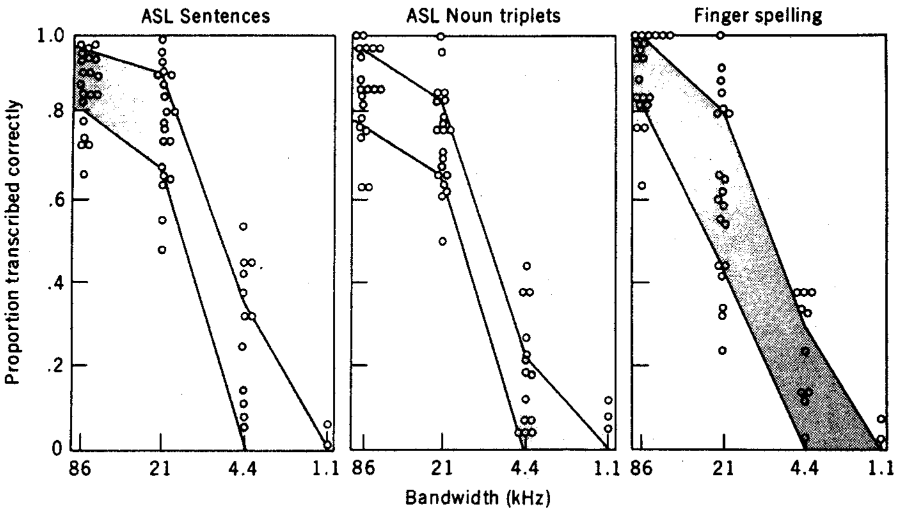

Fig. 2. The proportion of correctly transcribed items as a function of display bandwidth. Each data point represents the score of one of the 20 subjects in one condition. Subjects were excused from tests when it was obvious their scores would be near zero; only scores of tests actually taken are shown. The shaded area represents the performance of the middle 50 percent of subjects.

subjects to perfectly comprehend words produced at rates of three or four letters per second. The scatter of data points represents the large variation in language skill that exists among the deaf; individual measurements are reproducible.

Median ASL intelligibility at 21 kHz was about 90 percent of the control value. The loss at 21 kHz was somewhat greater for finger spelling. The best subjects (which include four of the five laboratory and university employees and just two or three of the others) were generally the youngest subjects in this sample: their ASL performance at 4.4 kHz was 40 to 50 percent of control. Their performance at 4.4 kHz is about what would be obtained with hearing subjects listening to voice communication over a 1.5-kHz (low-pass) channel (*15*). Thus, while a television picture may use more than 1000 times the bandwidth of a telephone line, visual transmission of ASL requires only a few times more bandwidth than voice communication.

The minimum bandwidth required for the video transmission of ASL by a raster scan is not known; the design of the present study yields only an upper bound. A more judicious choice of raster variables (fewer frames per second, more lines per frame, interlace, and so forth) would almost certainly reduce this upper bound substantially. The low bandwidth—the low information rate—makes possible more advanced picture coding by any of many schemes to further reduce the bandwidth.

Present video transmission systems thus use a much wider bandwidth than is required for transmission of ASL. By a sevenfold further reduction in the bandwidth required for ASL transmission or by a sevenfold increase in the bandwidth that can be carried on telephone facilities, the ASL-signing deaf population could—with appropriate video terminals—use our present telecommunication facilities for ASL communication.

GEORGE SPERLING

Department of Psychology,
New York University,
New York 10003

References and Notes

1. M. D. Fagen, Ed., *A History of Engineering and Science in the Bell System: The Early Years (1825–1925)* (Bell Laboratories, Murray Hill, N.J., 1975).
2. J. D. Schein and M. T. Delk, Jr., *The Deaf Population of the United States* (National Association of the Deaf, Silver Spring, Md., 1974).
3. H. Bornstein, in *Sign Language of the Deaf*, I. M. Schlesinger and L. Namir, Eds. (Academic Press, New York, 1978), pp. 333–361.
4. P. A. Bellefleur, *Volta Rev.* **78**, 107 (1976).
5. U. Bellugi and S. Fischer, *Cognition* **1**, 173 (1972); U. Bellugi and E. S. Klima, in *Deaf Children: Developmental Perspectives*, L. Liben, Ed. (Academic Press, New York, 1978), pp. 43–68.
6. D. E. Pearson, *J. R. Telev. Soc.* **16**, 6 (1978).
7. I. Dorros, *Bell System Tech. J.* **50**, 221 (1971); T. V. Crater, *ibid.*, p. 235.
8. G. A. Gerrard and J. E. Thompson, *Radio Electron. Eng.* **43**, 201 (1973).
9. Technical Staff, Bell Telephone Laboratories, *Transmission Systems for Communications* (Western Electric Company, Inc. Technical Publication, Winston-Salem, N.C., rev. ed. 4, 1971).
10. G. Sperling, in *Deaf Children: Developmental Perspectives*, L. Liben, Ed. (Academic Press, New York, 1978), pp. 103–114.
11. D. E. Pearson, *Transmission and Display of Pictorial Information* (Wiley, New York, 1975).
12. The recordings were made with a camera (Panasonic WV204P) having 4-mm and 8-mm lenses (Apollo) and with a cassette recorder (Sony VO-2800). The playback monitor was a 17-inch cathode-ray tube (Conrac QQA 17/N). For bandwidth calibration, vertical gratings that produced different numbers of cycles per centimeter on the monitor were recorded, and their modulation depth was measured with a microphotometer. Modulation depth diminished sharply to 50 percent of the maximum depth for coarse gratings at 2.0 ± 0.1 MHz. The sharp cutoff at 2 MHz was caused by the recorder; the camera and monitor easily passed 4 MHz. Recordings were made without interlace, producing 60 full pictures per second.
13. The number of horizontal raster lines in the four bandwidth conditions, respectively, was 79, 38, 17, and 9. (The full, noninterlaced television picture contained 262 lines.)
14. T. J. O'Rourke, *A Basic Course in Manual Communication* (National Association of the Deaf, Silver Spring, Md., ed. 2, 1973).
15. N. R. French and J. C. Steinberg, *J. Acoust. Soc. Am.* **19**, 90 (1947).
16. I thank E. Bloom, Jr., a deaf retired chemist, for acting as the signer for this study.

31 December 1979