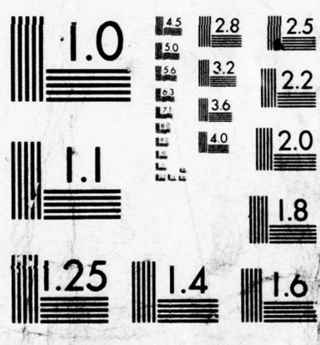
AD-A049 247 AIR FORCE HUMAN RESOURCES LAB BROOKS AFB TEX F/G 5/9
DISPLAY AND SPEECH DEVICES FOR SIMULATOR INSTRUCTOR/OPERATOR ST--ETC(U)
DEC 77 N F SCHWARTZ

UNCLASSIFIED AFHRL-TR-77-50 NL

| OF |

AD
A049 247

END
DATE
FILMED
2 — 78
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFHRL-TR-77-50

R FORCE

HUMAN RESOURCES

DISPLAY AND SPEECH DEVICES FOR SIMULATOR INSTRUCTOR/OPERATOR STATION APPLICATIONS

By

Noel F. Schwartz

ADVANCED SYSTEMS DIVISION
Wright-Patterson Air Force Base, Ohio 45433

December 1977
Final Report for Period June 1975 – June 1976

Approved for public release; distribution unlimited.

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE,TEXAS 78235

NOTICE

This final report was submitted by Advanced Systems Division, Air Force Human Resources Laboratory, Wright-Patterson Air Force Base, Ohio 45433, under project 6114, with HQ Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base, Texas 78235.

This technical report has been reviewed and is approved for publication.

GORDON A. ECKSTRAND, Director
Advanced Systems Division

DAN D. FULGHAM, Colonel, USAF
Commander

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (14) AFHRL-TR-77-50	2. GOVT ACCESSION NO.
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) DISPLAY AND SPEECH DEVICES FOR SIMULATOR INSTRUCTOR/OPERATOR STATION APPLICATIONS	5. TYPE OF REPORT & PERIOD COVERED (9) Final rept. June 1975 – June 1976
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Noel F. Schwartz	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Advanced Systems Division Air Force Human Resources Laboratory Wright-Patterson Air Force Base, Ohio 45433	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62205F (16) 61142001 (17) 20
11. CONTROLLING OFFICE NAME AND ADDRESS HQ Air Force Human Resources Laboratory (AFSC) Brooks Air Force Base, Texas 78235	12. REPORT DATE (11) December 1977
	13. NUMBER OF PAGES 18
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 18p.	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

display
instructor operation station
simulator
speech recognition
speech synthesis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The Air Force Human Resources Laboratory (AFHRL) has the responsibility for research and development of advanced simulation techniques, including more efficient and more effective Instructor Operator Stations (IOS) which would possibly use newly developed display devices and techniques and speech response/recognition devices.

This review was undertaken to become better acquainted with the state of the art of hardware devices which could be used for the IOSs of advanced aircraft training simulators and to provide some guidance in these devices to designers, specifiers and users of IOSs. Attention focused mainly on display devices and speech response/recognition devices.

A survey of technical literature concerning display devices, and speech synthesis and speech recognition

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

404 415

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

Item 20 Continued:

devices was accomplished and contacts were established with a number of manufacturers and developers of these devices to determine the latest developments and potential applications. Also, literature was searched for R&D related to the application of such devices.

Some of the merits and shortcomings of a number of display devices (i.e., cathode ray tubes (CRT) and alternative but similar devices) are discussed and descriptions of their operation are included. Speech interaction with computers is also discussed in a similar manner.

It is concluded that new display devices will not significantly impact the general design or utilization of the IOS. Advancement of speech recognition could have a significant impact, but development beyond present capabilities does not appear imminent.

ACCESSION for		
NTIS	White Section ☑	
DDC	Buff Section ☐	
UNANNOUNCED ☐		
JUSTIFICATION		
BY		
DISTRIBUTION/AVAILABILITY CODES		
Dist.	AVAIL. and/or SPECIAL	
A		

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

PREFACE

The review reported herein was performed during the period from June 1975 to June 1976. It supports the Air Force Human Resources Laboratory project 6114, Simulation Techniques for Air Force Training, for which Mr. Don R. Gum is Project Scientist and task 6114-20, Advanced Instructional Features, for which Ms. P. A. Knoop is task scientist.

The author wishes to thank Ms. Knoop for her valuable suggestions and assistance.

The author and the above mentioned personnel are with the Advanced Systems Division, Air Force Human Resources Laboratory, Wright-Patterson Air Force Base, Ohio.

TABLE OF CONTENTS

Preceding Page Blank - Not Filmed

DISPLAY AND SPEECH DEVICES FOR SIMULATOR INSTRUCTOR/OPERATOR STATION APPLICATIONS

I. INTRODUCTION

Purpose

The purpose of this report is to provide a compendium of information on display and speech devices to those involved in the design and specification of Instructor/Operator Stations (IOS). The report should also be of value to others concerned with general input and output of information to a human operator, especially when a simple alphanumeric character exchange may not provide the optimum format.

The orignial impetus for the survey on which the report is based was to acquire in-house expertise in display and voice response devices. This expertise is required to promote optimum utilization of these devices in future IOSs. It is also to be applied in determining requirements for needed R&D, either in the development of the devices themselves or in their use. The report makes no attempt to identify the nature of such R&D. This will depend on the nature of the application, whether one or more of the multitude of possible applications in an IOS or application in other training devices.

Background and Scope

IOSs have grown in size and complexity to keep pace with the monitoring and control requirements of increasingly complex modern multi-million-dollar training simulators. The IOS began its existence as a relatively simple station from which to control the simulator. It contained the necessary on-off switches and some repeater instruments (e.g., altitude and airspeed) for use by the instructor in monitoring the state of the simulated aircraft. As the complexity of the simulator grew, with the addition of motion base, visual system and increased complexity of aircraft weapon systems per se, the IOS also grew in complexity. The activation of the simulator with a digital computer and the inclusion of advanced instructional features also added to the complexity of the IOS. Some sort of general-purpose display became imperative. This display has traditionally been a cathrode ray tube (CRT), or rather a number of CRTs. Six or possibly more have been used at some instructor stations.

Since it seems likely that the successful future development of the IOS will rely to a great extent on the development and proper use of displays, CRTs or other, these devices were emphasized in this survey. Speech synthesis and speech recognition techniques and devices were also reviewed because speech is man's primary means of communication. Significant developments in speech synthesis and recognition could lead to radical developments in the design of instructor operator stations.

The survey of display devices and speech devices consisted of a search of several periodical indexes, such as "Applied Science and Technology Index" and "Computer Abstracts," for articles pertaining to these topics. Those articles of interest were reviewed and in many cases referenced other articles which were reviewed. The "Government Announcements and Report Index" (for the past several years) was also searched and reports of interest were reviewed, as were a number of reports listed in two Defense Documentation Center bibliographic searches. Also, a number of developers and manufacturers of display devices and speech response devices were contacted. These contacts were made in order to become familiar with the latest techniques and applications of these devices and to obtain technical literature covering these devices.

The devices and techniques reviewed are organized into two main categories for purposes of this report. The first category is "General-Purpose Computer Display Devices;" the second category is "Speech Input/Output Devices." A large part of the first category is devoted to CRT displays and techniques because of their broad and varied uses. The remainder of this category is divided into other display devices and 3D displays. Other display devices include plasma panels and liquid crystal displays. Speech input/output devices are divided into speech sysnthesis and speech recognition devices.

Short summaries have been added after the display section and the speech section to briefly summarize the state of the art. Also, a summary table has been provided at the end of the display section to provide an overview of the types of displays, available and pending, along with some of their characteristics.

The two main categories and summaries are followed by some concluding remarks and recommendations concerning the availability and applicability of some of the devices which are discussed.

Technical information has been minimized so the report may also be of use to nonengineers.

II. GENERAL-PURPOSE COMPUTER DISPLAY DEVICES

Introduction

The displays to be considered are general-purpose in the sense that various types of information in various formats can be displayed (alphanumeric, graphic, etc.) rather than the specific information presented by meter, alphanumeric panels or other special-purpose displays. The general-purpose display most commonly used in simulator IOS applications is the CRT. Typical of other devices used or proposed for general-purpose use are plasma panels and liquid crystal pictorial displays.

Although printers and plotters are display devices, they do not play the same sort of interactive role as the devices considered herein and are not discussed.

Some of the information required at the instructor operator station is of a type which is difficult to display effectively in alphanumeric form. Therefore, CRT terminals also having graphic display capabilities have been used in some instructor operator stations. Other uses of the CRT at the IOS have been to monitor televised scenes of the cockpit or to monitor the visual simulation scene or other visual displays which are presented to the pilot. Other devices have been or are being developed which may eventually replace the CRT in some or all of these applications.

Some of the advantages and disadvantages, and operating characteristics of these other devices and the CRT, will be discussed in this section.

CRT Display Devices

Since the CRT has been a very versatile device for displaying information, it will be given prime attention in this report. It can be used to present alphanumerics, graphic representations (both two-dimensional (2D) and three-dimensional (3D) perspective), and pictorial – all of this with or without color for further encoding or highlighting of information.

CRTs have a number of advantages and disadvantages. Among the advantages are: (a) they are widely used and understood and are made for a variety of applications, (b) they come in many sizes with a variety of phosphors for various color and persistence applications, (c) they have relatively high writing speeds and a corresponding quick change of display or dynamic display capability, (d) they are relatively inexpensive, and (e) they are the only devices capable of stroke-writing a nondiscretized graphic presentation. Some CRTs are discretized, but of the other display devices to be discussed all are discretized; that is, the display is composed of a large number of dots or discrete points.

Some disadvantages of CRTs are: (a) they are unwieldy, the depth characteristically being greater than the height or width of the display surface, (b) they are somewhat vulnerable to catastrophic failure; e.g., broken glass or burnt-out filament, (c) ordinary types must be refreshed 30 times or more per second; and (d) very high voltage, typically greater than 5,000 volts, is required to accelerate the electron beam.

Raster Scan vs. Stroke-Written Displays

CRT terminals and displays may be divided into those that are raster scan, also called TV scan, and those that are random scan, also variously known as stroke-written, calligraphic, direct write, or selective address.

In a raster scan system the beam starts at the top of the CRT screen and systematically "paints" horizontal lines one below the other until reaching the bottom of the screen, thus producing a raster. The intensity of the beam is varied as it paints to create a picture. In most applications a raster frame is composed of two fields with every other line being painted during one field and the interleaving alternate set being painted during the subsequent field. This technique reduces flicker by completing an alternate line raster at twice the frame rate, the frame rate normally being 30 Hz and the field rate 60 Hz.

The random scan or stroke-written display is so named because the beam can be commanded from any point on the screen to any other point on the screen in one stroke, a display thus being composed of hundreds or thousands of such strokes/or vectors as desired. For monochrome displays the same CRT may be used for either writing system, the difference being only in the technique used to control the electron beam. One technique deflects the beam to selected addresses or points on the face of the CRT; the other technique deflects it in a systematic, raster-producing fashion.

A stroke-written scan can produce a high quality graphic display within the limits of a display which can be composed of multitudes of lines or arcs. This high quality is obtained because the display is not discretized into raster lines or line elements as in a TV or raster scan display. The stroke-written display, however, does not lend itself to pictorial displays where background and shaded areas are required or where solid objects must be perceived in a relatively complex dynamic environment. In some applications a mini-raster technique is used to provide some shaded area capability for the stroke-written display. Stroke-written displays represent solid objects as "wire frame" structures, whereas the raster scan display can represent solids or opaque images as compositions of shaded or colored areas. Because of the quantization effects of raster scan systems, however, they have not quite reached the performance level of stroke-written systems for the generation of graphic-type displays. Raster displays, although requiring scan conversion to arrange the display content into the element and line format, offer several advantages over direct write or stroke-written systems. These advantages include: (a) a cost advantage for multiterminal systems, (b) uniformity for application of devices, and (c) the capability of mixing pictorial and computer-generated data, in addition to some advantages noted earlier. The efforts which have gone into image quality improvement for computed visual scene simulation have greatly reduced some of the quantization effects inherent in raster scan. Effects such as the line segmentation or stair-stepping that occurs on lines that are not vertical or horizontal have been reduced.

For raster scan the length of time to paint a display is constant and therefore unaffected by the content of the display. The general requirement that the display be flicker-free dictates refresh rates on the order of 30/sec. Thus, an upper limit exists on the number of strokes that a given calligraphic display can generate per image. The raster scan format, however, allows convenient paralleling of image generation hardware so that the upper limit on image complexity is determined by the resolution of the raster. The resolution of the raster is, in turn, dependent on the number of scan lines, typically 500 or 1,000 lines and the number of elements computed per scan line. If video is to be presented, resolution is also dependent on the bandpass of the video system. Resolution is also limited by and dependent on spot size although there are some applications where a slight defocusing might be beneficial (Bunker, 1973).

Types of Phosphors. CRTs are available with any one of a large number of phosphors having various color and persistence characteristics. At least 50 types of phosphors have been used. P4 is the white phosphor having medium persistence which is commonly used in black and white television (TV). P22, which is red, green or blue is the phosphor used in color television. Some other common colors are the familiar green of the oscilloscope trace and the somewhat less familiar long-persistence orange of low repetition rate radar displays. P50 and P51 are two of the new phosphors used for beam penetration color, to be discussed later.

Shadow Mask CRT. The shadow mask color CRT as used in color TV is available for raster scan applications. It is a discretized display, however, and the dot matrix nature of this device is somewhat objectionable at the close viewing distances that are normal to computer terminal applications. Some computer terminals are available which do use this CRT to provide full color capability.

Beam Penetration CRT. Limited color capability for stroke-written displays may be provided by a beam pentration CRT. This CRT has a multilayer screen consisting of two phosphors of different colors separated by a layer of transparent dielectric material. A number of colors, usually four, (red, orange, yellow and green) can be obtained by dynamically varying the pentration of the electron beam into the phosphor layers. The penetration is varied by varying the accelerating voltage on the CRT, usually over a range of 3,000 volts or more, depending on the type of CRT. A high brightness color display with over 1,500 lines resolution can be obtained with the beam pentration CRT. Another advantage is that no convergence circuitry is needed to superimpose the elements of a picture since only one electron gun is used rather than the three required in a shadow Mask CRT. A significant disadvantage is that the display content is reduced somewhat by the additional time which must be allotted to the switching of the high voltage which is necessary to produce color changes. Graphics terminals using this tube are commercially available for calligraphic color applications.

Storage CRTs. Another type of CRT for special applications is the direct view storage tube. One type retains an image up to about 1 hour without refreshing or until the image is erased. Because of the storage characteristic, it does not require continual use of buffer storage space as an ordinary display does. Selective erasure is not practical and it therefore cannot display dynamic imagery and there can be no changes of a picture part without the picture being erased. Due to these characteristics it has limited application as an interactive display device. Recently a technique has been developed to combine storage tube and refresh display techniques. One marketed terminal using this technique employs two internal computers to control the picture. Refresh graphics can be written at a rate of 5,400 vector-cm/sec and storage graphics at 13,500 vector-cm/sec. Vector-centimeters are the product of the number of strokes written by their lengths in cm. One limitation is that dynamics are limited to a portion of the display. The picture is flicker-free for up to 1,600 vector-cm or 800 vectors, whichever, comes first. To write more than this requires a slower refresh rate to allow sufficient writing time. Another type of direct view storage tube is the cathodochromic CRT. Recent advances have been made in this type of CRT by MIT. The cathodochromic CRT can store an image indefinitely and selective erase is possible; erasure, however, takes about 1 second.

Other Display Devices

Plasma Panel. One of the more successful competitors of the CRT has been the plasma panel which was developed in the mid-sixties. It is a dot matrix display device, a typical type having a display area about 9 by 9 inches with a dot matrix 512 by 512. This device consists of two quarter-inch plate glass substrates, each having parallel conductive electrodes applied to one surface. A thin layer of transparent dielectric material is placed over the electrodes, leaving their ends bare to allow connection to the control/drive circuitry. The two glass substrates are sealed together with the sets of electrodes orthogonal to one another to form the grid of the display matrix. A neon-based gas mixture fills the space between the sealed substrates. The completed panel is about 1/2-inch thick. When the proper voltage is applied across two electrodes, an electrical discharge is created through the gas at the intersection of electrodes which causes an emission of visible light. The display made up of the selected dots has the characteristic orange color of neon bulbs. The cells are written in an addressing technique similar to a core memory in which a single cell or row of cells can be addressed. The plasma panel is used for alphanumeric or graphic displays. The graphics, however, are subject to quantization effects as mentioned earlier such as the step effect on lines which are not vertical or horizontal. These displays have had wide application in the Plato computer terminals of the University of Illinois (Bitzer & Slottoro, 1967).

Some advantages of the plasma panel are: (a) a thin flat profile, 1/2-inch thick, (b) flicker-free performance, (c) inherent memory capability, (d) good shock and vibration characteristics, (e) good brightness and contrast ratio (60 foot lamberts and 25:1, respectively), (f) a transparent viewing area so rear projection through the display and ability to generate hard copies from either side are built-in features, (g) storage properties which allow interrogation of the display by the computer, and (h) high reliability and long life.

Some disadvantages are: (a) It is a monochrome (orange) display, (b) requires moderately high voltage compared to semiconductor circuitry (approx. 400 V) although not nearly as high as CRTs, (c) resolution is limited to dot matrix composition, (d) intensity control is limited, although some apparent area shading can be accomplished by varying the number of lighted dots per given area, and (e) write and erase time, 20 μsecs/cell, is too slow for some applications.

Sizes larger than 9" by 9" are available; however, there is some problem maintaining accurate spacing between the substrata for larger sizes. The cost of plasma panels is several times that of CRTs and may not be justified unless compactness and ruggedization are important considerations. Several standard sizes are commercially available. Several companies are using plasma panels in commercially available computer terminals. There has been some experimentation with three-color plasma panels (Hoehn & Martel, 1973).

Liquid Crystal Display. Another flat panel display device, still under development, is the liquid pictorial display which is a general-purpose outgrowth of the liquid crystal numeric displays used in some digital watches. It can be used for the presentation of graphic or pictorial information in addition to alphanumerics. A display is formed by controlling the relative reflectivity and hence the apparent brightness of each element of an array of elements. 1- by 1-inch arrays of 15,000 elements have been fabricated. A prototype developed in 1975 for the Air Force is made up of four of the 1-inch squares, each containing 10,000 elements. The display produces a black and white picture when fed from a TV camera or other video source. The liquid crystal display is a sandwich construction about 1/4-inch thick. It is fabricated by placing the appropriate liquid crystal material between a transparent conductive electrode and the large semiconductor chip consisting of metal oxide semiconductor (MOS) picture element drive circuitry, integrated on a substrate that forms the rear of the device. Shades of gray can be obtained because the relative reflectivity is proportional to scattering level, which is in turn proportioned to the applied potential.

Some advantages of the liquid crystal display are: (a) it will not wash out, even in direct sunlight, because it is dependent on reflected ambient light; the brighter the ambient light, the brighter the display; (b) very little power is required because the liquid crystal display modulates ambient light reflected through the crystal; (c) it has potentially high reliability; (d) it operates on low voltages, about 25 volts, as compared to several hundred for plasma panels and several thousand for CRTs; and (e) an important advantage is the response time which allows the display to be driven in real-time; it can, for example, be driven by conventional commercial television signals.

Some disadvantages are: (a) in darkness, some ambient light must be supplied; and (b) the size of the display is limited to the size of the MOS chip, about 2- by 2-inches, for currently available processing equipment. The display size of this type of equipment is expected to increase soon. It is expected that eventually the size of the display will be determined by techniques of assembling mosaic arrays of chips.

Larger panels 6″ by 6″ have been fabricated using a different fabrication process, however, resolution was only 20 lines per inch. (Brody, 1975b). Commercial availability of liquid crystal pictorial displays is not expected before late 1977.

Transistorized Viewing Screen. This device, still under development, is another flat panel display. A prototype was demonstrated in 1974 (Brody, 1975a). It is no thicker than a pane of glass. The glass is coated with layers of phosphor and microminiature thin film transistor circuitry. The prototype is a 6-inch-square panel capable of displaying alphanumerics, symbols, and lines with the appropriate electrical signals fed through its edges. A layer which is thinner than a coat of paint contains 36,000 electronic components. The thousands of sub-circuits are arranged in a matrix of dots. Activated sub-circuits cause the electroluminescent phosphor which comes in contact with them to glow. It is expected that screens will be rugged, reliable and inexpensive and require low power. Other advantages are that the brightness of each picture element can be varied to create shades of gray and color. Each element can be operated independently without activating other elements in the same row or column and picture elements do not require individual signal-carrying wires. Development is underway to improve resolution and to develop full color. Commercial availability is not likely before late 1978.

Miscellaneous Devices and Techniques. Some other techniques for generating displays are of the formatted alphanumeric type, in which seven or more bars or a dot matrix are selectively lighted or activated to form a number or letter. These will be only briefly mentioned because they are not normally used for general-purpose displays.

One very common device is the light-emitting diode. Arrangements of these are commonly used to form the displays of hand-held calculators and many digital wristwatches. They are also used in certain terminal applications. Of particular interest is a hand-held terminal which can display up to 20 alphanumeric characters. Two lines of 10 are developed while holding up to 100 lines in a display buffer for review when desired using a "scroll" switch. A 4 by 5 button keyboard allows input of the entire 128 character American Standard Code for Information Interchange (ASCII) set. Each button has a normal character and three alternates which can be selected by three switches operated with the holding (left) hand.

Another non-general-purpose display is the electrochromic display. This display is produced by precipitating a compound out of solution onto a cathode. Although no energy is required for maintenance of the display and very little for switching the display, the device presently operates too slowly for many applications. Research is underway to further develop and improve these devices.

3D Displays

3D displays offer the possibility of presenting some forms of complex data to an operator in a way that enhances the operator's ability to assimilate such data.

Presently there are at least four general approaches to the problem of presenting multi-dimensional information. They are: (a) coding in 2-D presentations, (b) perspective and other monocular cues in 2-D presentation, (c) stereoscopic 2-D systems, and (d) volumetric (3-D) devices (Flackbert et al., 1972).

The first of the techniques mentioned previously adds third-dimensional information to a 2D display by coding the third dimension information into graphical, color or alphanumeric form. This technique has been used successfully in some applications; however, display clutter may offset the advantages if not used cautiously.

Monocular cues such as size or shading may be used to produce the illusion of depth on a 2D display. Perspective, such as the display of an object as an isometric projection, is especially useful in producing the illusion of depth. Some of the problems associated with this technqiue; such as the generation of complex shapes and the elimination of hidden lines, have benefitted from recent work in computer image generation.

Stereoscopic viewing of separate 2D displays, one for each eye, employs the use of binocular disparity as an additional cue to enhance the illusion of 3D. Although this technique offers a realistic illusion of depth, it generally suffers the basic deficiencies of single-vantage viewing point and requires the wearing of special viewing devices. Recent computer-generated displays have somewhat relieved the constraint of single-vantage-point viewing by measuring the position of attitude of the observer's head and changing the displayed object accordingly.

A volumetric display is the only one of the techniques that presents depth in a real rather than synthetic way. All of the optical cues are used – binocular disparity, perspective, size and focus. One such volumetric display employs two beams of electromagnetic energy intersecting at a point in space. When the proper medium is employed, visible fluorescence results at the beam intersection; thus points of light can be created in three-dimensional space. This technique, called Sequentially Excited Fluorescence (SEF) was confirmed by laboratory experiments conducted by Batelle (Flackbert et al., 1972). Full development by 1978 was predicted.

Summary of Display Devices

There is no substitute for the CRT for non-discretized displays or for full color displays. Also they are versatile and inexpensive, but somewhat fragile.

Plasma panels can substitute for CRTs in some applications. Their flat panel configuration has the advantage of space saving and they are somewhat more rugged than CRTs. They have a unique rear projection capability. Commercially available types, however, are monochrome and limited in size to about 9 by 9 inches.

Liquid crystal pictorial displays and transistorized viewing screens are still under development. They promise no unique advantages over CRTs for most applications other than the advantage of flat panel configuration and low voltage operation.

Computer techniques can add 3D perspective to 2D displays. True 3D volumetric displays are still under development.

Hand-held terminals with alphanumeric display are not commercially available. See Table 1 for a summary of display devices.

III. SPEECH INPUT/OUTPUT DEVICES

Introduction

The effective use of the IOS of modern aircraft simulators requires extensive interaction between man and computer. This interaction is usually through manual and visual channels. Speech is man's primary means of communications. Therefore, man-to-computer and computer-to-man communication by speech is very appealing and continues to receive a good deal of research effort. However, the value and even the practicality of speech interaction with machines is one of the interesting unresolved issues in computer science (Hill, 1972).

Some of the advantages summarized in the above reference are: (a) speech is more natural,

Table 1. Display Summary Table

	Typical Use	Color	Display Size	Discretized	Response	Required Refresh Rate[a]	Selective Erase	Commercially Available	Major Advantage/ Disadvantage
Conventional CRT	Black & White TV	Monochrome	LG. ($>$ 21″)	No	Fast	High	No	Yes	Versatile/ Depth Size
Shadow Mask CRT	Color TV	Full	Lg.	Yes	Fast	High	No	Yes	Full Color/ Discretized
Beam Penetration CRT	Color Computer Terminal, Stroke Written	Multi (4)	Lg.	No	Fast	High	No	Yes	Stroke Color/ Limited Colors
Storage CRT	Special Oscilloscope	Monochrome	Lg.	No	Slow	Low	No	Yes	Low Refresh/ Slow Response
Cathodochromic CRT	Experimental	Monochrome	Lg.	No	Slow	Low	Yes	No	Low Refresh/ Experimental
Plasma Panel	Computer Terminal	Monochrome	Med (9″)	Yes	Med.	Low	Yes	Yes	Flat Panel/ Monochrome
Liquid Crystal Pictorial Display	Experimental	Monochrome	DM(2″)	Yes	Fast	Low	Yes	No	No Washout/ Size
Transistorized Screen	Experimental	Monochrome	Med(6″)	Yes	Med.	Low	Yes	No	Flat Panel/ Experimental

[a]Also may depend on application.

more convenient and interferes less with other activity than other channels; (b) speech provides an extra channel for multi-mode communication (for example, allowing control and data functions to be separated in a computer-based text editing system, controlling the editor by voice and entering text data on a keyboard; (c) speech is very suitable for alert or "break-in" messages and responses (designed, for example, to interrupt an interaction); (d) speech uses a minimum of panel space; (e) speech is compatible with normal means of communication (for example the inexpensive and ubiquitous telephone system); (f) speech is independent of factors affecting sight and reach; and (g) speech is easily monitored by a third party (for example, supervision of data gathering using interactive terminals for medical screening). It is noted that all of these apply equally well to either input or output for machines. The disadvantages are similarly summarized in the same reference. A speech interface: (a) leaves no permanent record (awkward for scanning or verification, though simple audio recording is some help and for pass-words this could count as an advantage) (b) may be unreliable due to crosstalk (the cocktail party type of problem), (c) could give a false sense of the power and understanding possessed by the machine, and (d) at least for input, may prove more expensive than other forms of input. Additional disadvantages not cited in the above reference are: (a) speech is relatively slow for information presentation, and (b) it is not practical for the presentation of graphical or tabular information.

Some problems associated with the development of speech communication are interference by ambient noise, the varition of voice and speech characteristics with individual speakers, and the present lack of complete knowledge of acoustic, linguistic and semantic aspects of computer processing of speech as well as the processing requirements and cost associated with the latter (Turn, 1974).

Speech Synthesis

There are a variety of well-known techniques for generating speech output from machines. These techniques range from pre-recording natural voice words or messages to synthesis of speech derived from vocal track resonances and the various other characteristics of the human speech – sound generation system. One straightforward approach is to voice record one word per channel on a multitrack device such as a drum and have the computer select the words in the proper sequence to form a message. Although all the words are sensed at the same time, the switching logic insures that only one word channel is selected and actually output during each word cycle. The recordings, which can be either magnetic or optical, are subject to the usual mechanical wear and tear and maintenance problems associated with mechanical devices when played over and over. It has been noted also (Flanagan, 1972) that for good results such messages can be used only in the context in which they were recorded. A device of this type has been used in an advanced instructor station.

Another approach to speech synthesis has been to break words into phonetic segments and then digitally store each segment in some form of memory. The voice is recreated by accessing the memory through the appropriate programming to reconstruct the phonetic segments into words. Although a rather choppy reproduction without natural qualities has typically resulted from this approach, it has been used and found adequate for certain applications. If sufficient phonemes, 30 or so, are stored, an unlimited vocabulary can be synthesized.

One approach has been reported by one company which digitizes and stores whole words in solid-state read-only memory and is said to retain natural voice qualities and inflections. A proprietary approach is used involving the complete analysis of plotted audio waveforms. The conversion of the analog audio signal of a word into a digital signal required as few as 8,000 bits of memory storage rather than the usual 40,000 bits. Other techniques which put together signals in various frequency channels can produce intelligible speech with as little as 2,000 bits of memory per second of speech.

It has been noted, however, by Flanagan (1972) that if the computer is to speak with a large sophisticated vocabulary and if it is expected to use this vocabulary to form a wide variety of messages and contexts, the simple technqiue of prerecorded natural speech is ruled out. Economical storage of large amounts of speech data in a form flexible enough to generate arbitrary messages implies a speech synthesis approach, an approach, that is, which would model the speech process and use control functions and data obtained from natural speech utterances or from programmed knowledge of the speech process to synthesize speech. Speech synthesis devices are marketed by a dozen or more firms, including major computer manufacturers such as IBM and Honeywell. Many have used the multitrack, prerecorded, one word per track technique using conventional magnetic or optical recording. More recently devices are being marketed using the technique of storing words or phonetic segments in solid-state memory. These devices are quite reasonable in price.

Speech Recognition

Although several manufacturers now make voice recognition equipment, it has not reached the stage of practicality that speech synthesizers have. It is noted by Hill (1972) that the difficulty in achieving real progress in automatic speech recognition lies in the fact that true speech recognition by machine would require the automation of two human abilities which are not well understood – auditory perception and understanding. The traditional steps involved in speech recognition as given by Hill are: (a) transduction of the acoustic pressure waveform into machine-sensible form (for example, a microphone produces an electrical waveform related to the original pressure waveform); (b) analysis of this waveform to provide some measurement space (frequency measures, binary measures related to complex components, etc.); (c) transformation of the points in the measurement space to provide a decision feature space for recognizable units (very often phonemes are chosen as the recognizable units because these form the basis for conventional linguistic descriptions of speech utterances, are finite in number – of order 40 in fact – and bear a reasonable resemblance to normal written language); (d) the making of a decision as to which recognizable unit occurred (this terminates "acoustic recognition," with the provision noted above concerning interaction between states); (e) if the units recognized are phoneme-like, there follows a stage of recognizing words and/or phrases on the basis of "noisy" phoneme strings; (f) the recognition of the syntactic structure of the given string of words as now determined, which may include incorrectly recognized and/or omitted words; (g) the extraction of the semantic content, or meaning of the message on the basis of the words and the syntactic relationships discovered; and (h) the execution of some action on the basis of the meaning determined. The following description characterizes how one company accomplishes speech recognition (Herscher, 1975):

> An isolated word recognition system identified a spoken word by measuring acoustic characteristics of the word and comparing these characteristics of the word with those of reference words: stored in memory. The reference word characteristics are input beforehand by the same speaker who will be using the system. A set of 32 features is used to characterize a word. These features are of two types; five broad class features and 27 phonetic event features. The five class features are: (a) Vowel/Vowel-like – which occurs for all vowels and vowel-like consonants, (b) Long Pause – which occurs for all pauses greater than 100 msec, (c) Short Pause – which occurs for all pauses less then 100 msec, (d) Unvoiced Noise – like consonant – which occurs for all unvoiced fiction – produced and unvoiced stop consonants, and (e) Burst – which occurs for the abrupt onset of energy for some phonetic transitions; i.e., stop consonant-to-vowel. The 27 phonetic event features represent measurements corresponding to phoneme-like occurrences. The word recognition system consists of three subsystems: Preprocessor, feature extractor, and classifier. Both the preprocessor and feature extractor functions are hard-wired. The classifier function is performed by software in a minicomputer. For each spoken word, the 32 encoded features and their time of occurrence are stored in a short-term memory. When the end of an utterance is detected by the feature – extractor logic, the duration of the word is divided into 16 time segments and the features of the word are reconstructed into a normalized time base. Pattern matching logic subsequently compares these feature occurrence patterns to the stored reference patterns for the various vocabulary words and determines the "best fit" for a word decision. There are 512 bits of information (32 features mapped into 16 time segments) required to store the feature map of an utterance or reference pattern.
>
> During the training mode, the system automatically extracts a time-normalized feature map each time the speaker repeats a given word. A consistent matrix of feature occurrences (between repetitions of the word) is required before the word features are stored in the memory. In the operational mode, each word spoken into the system is processed in the same way as were the training words – that is, word features are extracted, digitized, and time normalized.
>
> The resultant word matrix is then digitally compared to each stored reference matrix. The stored reference word producing the highest overall match is then selected by the system and an output decision is made.

Current voice recognition systems are based on a highly constrained manner of speaking. For good recognition accuracy, commands, i.e., words, must be separated by a pause of about 200 msec. and vocabulary elements must be selected to eliminate easily confused commands. Also the system must be trained to a particular speaker. This is accomplished by having a speaker repeat (usually 5 to 10 times) each word to be recognized by the system. It can be trained for additional speakers and the results stored in mass storage where facilities permit. With the above factors optimized a recognition accuracy of 95 to 100% can result.

In the operational mode, each word spoken into the system is processed in a manner similar to the training procedure wherein acoustic features are extracted, digitized, and time-normalized. The resultant test word matrix then is compared digitally to each stored matrix. One means of expanding the vocabulary, forming short phrases, and speeding the response is by narrowing the comparison search using a structured vocabulary format. The first word must be from a list of select words. A second word is selected from a group of words determined by the selection of the first word. The second word, in turn, determines from which group of words a third word of the phrase is selected.

Word comparisons are limited by computer speed to about 30 to 40 due to the complex processing which must be accomplished in near real-time. Word recognition can, however, easily be extended to 90 or more by making use of the 30-phrase, 3-word format. This structual vocabulary technique has, in fact, been extended to increase recognition vocabulary to over 300. A minimal speech recognition system consists of: (a) a preprocessor which reduces every word, regardless of length to a bit pattern or matrix; (b) a minicomputer consisting of processor with 16,000 bytes of memory which compares the spoken word matrix to each stored reference matrix; and (c) a display for verification of the spoken commands. Larger systems may also include speech synthesis units and standard peripherals including a disk which is utilized to store a number of different user reference patterns and the operational software.

At this time, only two or three firms are known to be marketing speech recognition systems. Such systems are available with output in the same format and code as that of a standard key board terminal. Some firms active in the development of speech recognition are: Threshold Technology Inc. of Delton, NJ; Scope Inc. of Reston, VA; and Dialog Systems, Inc. of Cambridge, MA.

Summary of Speech Input/ Output Devices

Speech synthesis devices are available in a variety of types and prices. Solid-state devices using phonemes stored in read-only memory are available at very reasonable prices. They are lacking in natural voice qualities, but are readily understandable and capable of versatile speed output. Higher quality speech can be attained by assembling prerecorded words, but this system does not have the versatility of vocabulary obtained from assembling phonemes.

Speech recognition is much more complex than speech synthesis. The speech recognition device must be trained for each speaker. Speech must be in a halting, unnatural style and using only the vocabulary to which the device has been trained. These devices are therefore mostly successful in applications where other forms of input are impractical.

IV. CONCLUSIONS

A knowledge of the hardware options (which are available for the implementation of an IOS) helps to promote an optimum hardware design of the IOS. Various options may be considered and trade-offs between cost and effectiveness and other factors selected.

A number of the devices described are in the prototype stage or are not well proven and others may be too expensive or not applicable for other reasons. This report should, however, serve as an introduction to some of the developments which may be proposed or expected.

There are a great number of applications besides IOSs driving the development of speech and display devices. The state of the art is changing quite rapidly as a consequence.

The display devices which are developing as alternatives to CRTs are not expected to significantly impact the general design or utilization of the IOS. However, these devices may influence such factors as size, ruggedness, and cost.

Speech synthesis has been used in an IOS and its continued use is likely to be beneficial in certain applications. Speech recognition, when sufficiently developed, offers great potential for advancement of the IOS. At the present stage of development, however, its benefit to an IOS has yet to be demonstrated.

It has been noted that every problem of presentation of information arises from the needs of a human operator. A thorough understanding of these needs in a particular application such as IOS is essential to good system design. The same information, for example, can be displayed in a vast number of ways from tables of binary numbers to some type of graphic presentation. The binary form, obviously is very ineffective for most applications. The prime consideration should be how displays and other components of the IOS are used, rather than what devices are available. It is beyond the scope of this report to delve into the particular needs of human operators or various other human factors. It is important to note, however, that it does make an important difference how information is portrayed and formatted. Research is needed to develop objective techniques to improvement of the IOS.

REFERENCES

Bitzer, D.L., & Slottoro, H.G. The plasma panel – a new device for direct display of graphics. *Emerging Concepts in Computer Graphics*, University of Illinois Conference, 1967, 13–28.

Brody, T.P. A 6 x 6 in. 20-lpi electroluminescent dispaly panel. *IEEE Transactions on Electron Devices*, 1975, ED-22, 739. (a)

Brody, T.P. Large scale integration for display screens. *IEEE Transactions on Consumer Electronics*, August 1975, CE-21(3), 260–288. (b)

Bunker, W.M. Image quality improvement in computed visual scene simulation. *Proceedings of the 6th NTEC/Industry Conference*, 1973, 96–124.

Flackbert, C., et al. *3D Volumetric display study*. AD-753 438. Scottsdale, AZ: Motorola Incorporated, Government Electronics Division, December 1972.

Flanagan, J.L. Voices of men and machines. *The Journal of the Acoustical Society of America*. May 1972, 51(5), 1375–1387.

Herscher, M.B., & Cox, R.B. Talking to your machine. *Automation*, March 1975, 65–69.

Hill, D.R. An abbreviated guide to planning for speech interaction with machines: The state of the art. *Interm. Journal of Man-Machine Studies* 1972, (Vol. 4), 383–410.

Hoehn, H.J., & Martel, R.A. Recent developments in three-color plasma display panels. *IEEE Transactions on Electron Devices*, 1973, ED-20, 1078–1081.

Turn, R. Speech as a man-computer communication channel. AD-783 319. Santa Monica, CA: Rand Corporation, January 1974.

BIBLIOGRAPHY

Faconti, V., & Epps, R. *Advanced simulation in undergraduate pilot training: Automatic instructional system.* AFHRL-TR-75-59(IV), AD-A017 165. Wright-Patterson AFB, OH: Air Force Human Resources Laboratory, October 1975.

Flanagan, J.L., Ishizaka, K., & Shipley, K.L. Synthesis of speech from a dynamic model of the vocal cords and vocal tract. *The Bell System Technical Journal,* March 1975, 54(3), 485–506.

Foley, J.D. The art of natural graphic man-macine conversation. *Proceedings of the IEEE,* April 1974, 62(4), 462–471.

Glenn, J.W. Machines you can talk to. *Machine Design,* May 1, 1975, 72–75.

Kinsella, K.J., & Matthews, A.J. Using interactive graphics for fighter pilot training. *Information Display,* March/April 1972, 9(2), 15–20.

Knoop, P.A. Advanced instructional provisions and automated performance measurement. *Human Factors,* December 1973, 15(6), 583–597.

Licklider, J.C.R. Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics,* 1960, HFE-1, 4–11.

Marill, T. Automatic recognition of speech. *IRE Transactions on Human Factors in Electronics,* March 1961, HFE-2(1), 34–38.

Newman, W.M., & Sproull, R.F. An approach to graphics system design. *Proceedings of the IEEE,* April 1974, 62(4), 471–482.

Newman, W.M., & Sproul, R.F. *Principles of interactive computer graphics.* New York: McGraw-Hill, 1973.

Plott, H.H., Jr., Irwin, D.J., & Pinson, L.S. A real-time stereoscopic small-computer graphics display system. *IEEE Transactions on Systems, Man, & Cybernetics,* September 1925, SMC-5(5), 527–533.

Potts, J. Computer graphics – whence and hence. *Computer and Graphics,* September 1975, 1(2/3), 137–156.

Saltz, W.L. Survey of hardware R&D for computer displays. *Computer Design,* May 1972, 89–93.

Smode, A.E. Recent developments in instructor station design and utilization for flight simulators. *Human Factors,* February 1974, 16(1), 1–18.

Van Dam, A., Stabler, G.M., & Harrington, R.J. Intelligent satellites for interactive graphics. *Proceedings of the IEEE,* April 1974, 62(4), 483–492.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1977–771-122/81