Automated therapy for nonspeaking autistic children*

by DAVID CANFIELD SMITH, MALCOLM C. NEWEY and KENNETH MARK COLBY

Stanford University
Stanford, California

CHARACTERISTICS OF AUTISTIC CHILDREN

Earlier publications described our computer method for stimulating language development in nonspeaking children, sketched several case histories (Colby 1968⁴), and gave statistical evidence that our high rate of success (71 percent) was due to our treatment method (Colby and Smith 1970⁵). This paper presents some proposals for making the method more widely available.

The term “Childhood autism” refers to a psychiatric category of mental disorders occurring in children. Though the category is somewhat unreliable, there is a set of characteristics generally associated with autism. Some of the characteristics of autistic children that influenced our approach and our choice of them as subjects are described briefly. More complete descriptions are provided by (Rimland 1964⁴), (Wing 1966¹⁴), (Ornitz and Ritvo 1968¹⁰), (Colby 1968⁴, Colby and Smith 1970⁵).

An autistic infant’s early disinterest in play or being picked up develops into a clear unresponsiveness to people during his second and third years. He avoids contact with other people, sometimes running and hiding from strangers. He does not respond when spoken to, and may be diagnosed as deaf. He may be fascinated by spinning objects and machines, such as phonographs, washing machines, light switches, etc. Far from being mentally retarded, an autistic child may be more adept than normal children at music and mathematics, and may have a better memory.

But the most characteristic attribute is his inability to acquire language. An autistic child has great difficulty understanding the relation between human vocal sounds and meaning. Those that do develop some speech have trouble advancing beyond nouns designating concrete objects to pronouns and verbs with their inherent abstractions. All studies agree that the prognosis for an autistic child is closely correlated with speech development; an autistic child who has not developed speech by the age of five has in the past almost invariably been institutionalized for the rest of his life. If a child has speech, therapy for his other disorders has a far greater chance of success; indeed, absence of language may in itself account for many of his other shortcomings. Having speech, he may begin to play with other children, to go to school, and to build a model of the world that is no longer devoid of concepts transmitted through language.

Though autism is admittedly an imprecise category, the best estimates currently available indicate the incidence of autism to be 4-5 per 10,000 children, about the same incidence as deafness and blindness in children.

The main thrust of our effort, then, is to stimulate language development using a method which is acceptable to an autistic child on his own terms. We give him a chance to play with a machine which happens to be a linguistic entity.

A COMPUTER METHOD FOR TREATING AUTISTIC CHILDREN

While our eventual goal is to implement our method on a system with modest or no computer support, our initial approach was dictated by the resources available to us and by our need for flexibility and modifiability. For these reasons we used the large computer system at the Stanford Artificial Intelligence Project, Stanford University. That system is a time-shared...
DEC PDP-6/PDP-10 combination (DEC 1968). The PDP-6 is primarily used as an I/O driver at Stanford; it controls digital-analogue and analogue-digital converters, among other input/output devices. The faster PDP-10 serves as the central processor. Actually a time-sharing system is not particularly suited to our method, because the need for rapid responses of long duration imposes real-time constraints which are incompatible with the philosophy of time-sharing. In fact, our need to output continuous sound of up to three seconds duration has not yet been implemented to our complete satisfaction.

Our goals of a flexible, easily-modifiable system with a minimum of programming effort initially led us to choose a high-level programming language MLISP (Smith 1970), a dialect of LISP. Admittedly this was like using a sledge hammer to kill a fly, but MLISP enabled us to concentrate most of our early effort on the program itself and not on developing a programming system. Subsequently, our entire system was rewritten (by Malcolm C. Newey) in machine language. It was decided to invest the additional programming effort to gain the greater efficiency of machine language because the developmental phase of many of the games had been completed, and they are no longer being modified. In addition to the basic program itself, we and others at Stanford have developed a set of programs and program modules that facilitate the construction and modification of games. Among these utility programs are a drawing program, with which figures and animation can be easily drawn on a display using a light pen, and a sound recording program with which sound of up to three seconds duration may be recorded, digitized, and stored on the disk.

For hardware we use a computer-controlled Information International Inc. (III) graphics display with a screen about 2' x 2' in size that is refreshed 30 times per second. The need for rapid response, including animation, argued against the economy of a storage-tube display. At present, input is by an alphanumeric keyboard resembling a typewriter keyboard; the child presses a key, and a response is generated. We are currently investigating the possibility of light-pen interaction, so that the child can draw his own figures, and of microphone interaction, so that the computer would be activated by a voice signal rather than a keyboard signal. An associated, rapid-response auditory display is generated by randomly accessed, digitized records of voice, animal and mechanical sounds. The sounds are recorded using one of the utility programs, digitized at 20KC, and stored on secondary storage (IBM 2314 disk). About 1000 audio-visual experiences are available in the currently used set of games, requiring nearly 5 million words of disk storage. When a sound is to be played, it is retrieved from the disk, passed through a digital-to-analogue converter, and output through a speaker system. The high recording rate (20KC) was arbitrary; we wanted good clear fidelity, but a lower frequency (10KC or 5KC) could be used since we tend to favor low, vibrant sounds.

Thus we utilize at Stanford a PDP-10, a III display with keyboard, an IBM 2314 disk drive with 5 million words of secondary storage, and a D/A converter with a speaker system. Clearly these heavy requirements make the widespread replication of our total system impossible. However, we have anticipated eventual implementations of the method on small machines, and even on non-computer-controlled devices (as discussed below), by a modular construction of our programs. The sound and display files are organized into separate and distinct "games". Each game emphasizes a particular aspect of linguistic development. In most cases they do not share files; each game may be separately loaded into the computer much the same (conceptually) as tape cassettes may be loaded into a tape recorder. (In fact, that is one possible implementation of a non-computer device.)

The different game modules may be summoned by the sitter, a non-professional adult who is present at all therapy sessions. The following descriptions of the games are from an earlier paper (Colby and Smith 1970). The descriptions are still valid because, as was mentioned earlier, these games are in their final form. Many games continue to be developed, but their philosophy is similar to that of the games presented here. These games are better described by the slides and recordings which will accompany presentation of this paper.

Game 4 is especially worth noting in view of the autistic child's impaired affective relationships with other people. Elements of compassion are introduced obliquely in several of the frames. For example, in one series a cartoon of a large ice cream cone is accompanied by sounds of loud, happy slurping, and the word SLURP! Then the top scoop falls off (PLOP!), and there is a groan of dismay. The sad faces of many children viewing this reveal clear communication of a shared emotional experience.

Game 1: Letters and numbers. The letter, number or symbol of the pressed key is displayed accompanied by a male or female voice pronouncing the appropriate sound. Many of the letters are shown in different sizes and shapes, and a few are drawn in motion automatically on the video screen.

Game 2: Words. Single words appear on the screen in response to pressing a key. The words consist of...
verbs and adjectives involving affect and motion, e.g., 'laugh', 'jump', and 'angry.' The affect terms are not only stated but enacted, i.e., the voice laughs while saying 'laugh.' All the major positive and negative affects are referred to in an attempt to exercise the child's affective repertoire. For example, the word 'kill' is spoken very belligerently, and the combination 'XYZ' is accompanied by a contemptuous Bronx cheer.

Game 3: Phrases. A letter or number is displayed in response to the corresponding key being pushed. Following the symbol appears a word or a phrase containing it one or more times, with arrows pointing to these occurrences. A male or female voice states the word or phrase. The words and phrases are selected on the basis of common occurrences in the child's life (e.g., 'funny', 'upstairs') and because of their absurdity (e.g., 'fried filet from Farnsworth', 'lambs go baaaaaaal').

Game 4: Cartoon pictures. This game consists of displayed phrases and pictures along with human, animal, machine and other sounds such as music. The figures consist mainly of animals and machines, with very few humans. Crude motion is achieved by moving the animals' legs or moving a vehicle across the screen. The pictures and sounds are intended to be ludicrous, absurd and amusing to children.

Game 5: Phrase completion. A phrase or sentence consisting of several words appears on the screen in response to a key; e.g., pressing 'J' produces 'the cow jumps over the moon.' A male voice expresses the sentence over and over until the child stops it by pressing any other key. When 'J' is pressed again the same thing happens, but on the third occurrence, while the full sentence appears, the voice omits part of it, saying e.g., 'the cow jumps over the ---.' The intent is for the child to fill in the omitted word or phrase, which is omitted one of three times.

Game 6: Word construction. Here the letters of the alphabet appear in a square around the edge of the video display. On pressing a key, a trumpet sounds with the Los Angeles Rams 'Charge!' call, a female voice says, 'Here comes the word,' a male voice states the word, e.g., 'eat,' and another male voice pronounces the letters, which one by one light up and move to the center of the screen to form the word. The word is again stated when the spelling is completed. The game demonstrates how words are put together letter by letter.

Game 7: Phrase construction. The alphabet is displayed at the bottom of the screen. In response to pressing a key, letters move to the top of the screen, following individual paths, to form a phrase, one word at a time. The effect is of chaotic motion resolving itself at the end into a coherent word.

Game 8: Typewriter. The video portion of this game consists of the symbols of the keys pressed appearing in lines from left to right until the screen fills up, whence they are all erased and a new sequence begun. The audio portion involves a male voice offering comments, requests and questions regarding the displayed symbols—e.g., 'Is this really eight?,' 'where is O?,' 'see the four?,' 'now find I.'

Game 9: Spelling. The child types freely. If, by chance or design, he spells any of the words which occur in any of the games, the rest of the screen is erased temporarily leaving the word he has spelled centered in the screen. The appropriate sound is then played.

We are continuing to experiment with new games in an effort to build a library of the most successful approaches to stimulating language.

THE RATIONALE FOR AUTOMATED THERAPY

Colby has already expressed our initial motivation for using a computer in automated therapy (Colby 1968, p. 642):

My group's interest in a computer-based method for developing language in non-speaking disturbed children derived from several sources. First, we were interested in the general problem of using computers in the problems of psychiatry, as for example through computer simulation of belief systems (Colby 1967a) (additional references: (Colby and Smith 1969), (Colby et al. 1971) and man-machine dialogues (Colby and Enea 1967b)). Second, the work of Suppes (Suppes 1966) and Moore (Moore 1963) indicates that normal children learn reading, writing, set theory, and arithmetic rapidly and enjoyably by means of computer-controlled keyboards and displays. Third, the observation of many workers regarding the great preoccupation of some disturbed children with mechanical objects which they can manipulate and control is impressive. Since language acquisition in a normal child results from interactions with people (relations which disturbed children find difficult), perhaps nonspeaking disturbed children would find a machine such as a computer-controlled keyboard and display a more acceptable source for linguistic interaction. Hence, an effort was made to take advantage of a child's fascination with machines by providing him...
with a speaking and writing machine to play with. Instead of a person controlling a child, the child can control the machine, making it talk and display symbols at his will.

Language is often described as used for expression and as an instrument for social influence. But during normal language acquisition, it is also used by children as a toy. The method used in the present studies offered each child a means of playing with language. The hunch that children might enjoy this activity was further supported by some preliminary experience with normal children who delighted in the play and whose speech was greatly excited by it during and after the sessions. If a nonspeaking disturbed child could become interested in this sort of play and begin to enjoy developing language as play rather than work, the hope was that he would transfer his use of language from a computer context to other social contexts. If a disturbed child talks, a greater chance exists for understanding what troubles him and for helping him.

In addition to eliminating the fear of interpersonal relationships possessed by many autistic children, the computer has other intrinsic advantages over a human therapist. Some children's linguistic difficulties stem from an inability to abstract from complex and changing situations. They are baffled by vocal sounds coming from people. Even the most well-intentioned human therapist will inject variety into the learning situation, thereby including one more variable with which the child must deal. The response of the machine is an absolute constant which forms a firm base for the child to build upon. We frequently observe children striking the same key over and over, and listening intently to the sound. "Adults do not tolerate repetition well, and when a child strikes the same key for twenty minutes, it takes great control not to interfere. We know from experience that if a sitter tries to stop repetitions, the child will resist the interference and refuse to play further. And he is right." (Colby and Smith 19700)

Has a nonspeaking autistic child lost hope of understanding and using language? Has he given up so that nothing seems worth the effort or risk? If so, we get him to try again, to rekindle the hope by providing him a linguistic experience requiring little effort and at which he cannot fail. For a small amount of effort in pressing a key, he produces a large effect; and there is no risk of failure because something always happens, and he is never corrected for being wrong about what happens. He is free to select those symbols he wishes to manipulate. He is free to exercise his curiosity, to explore, to try to make sense out of the games. The experience is designed to be fun, non-serious, not like school (BLEAGGH!). The program is full of foolishness, nonsense and absurdities intended to delight a child and to elicit glee and exuberance. Only personal observation of the children playing with the display can convince adults of how pleasurable it can be. (Colby and Smith 19706)

The strongest argument, though, for our method is the favorable results we have achieved. Other treatment methods for nonspeaking autistic children report an improvement rate of 3 percent-57 percent (Bettelheim 1967). To date (Nov. 1971) we have treated 21 cases over the past four years. We have estimated 15 (71 percent) to have improved in language development. We classify a child as nonspeaking if he is mute or utters only a word or two per month or year. He will also generally utter other sounds that are unintelligible. (However, all but one of our children have shown some language comprehension.) We rate a child improved if he moves from this state to volunteering appropriate intelligible words and phrases in social communication. The speech must be non-echolalic and be initiated by the child with the intention of communicating information. Though the enunciation need not be perfect, it must be sufficiently good so that the child can make practical use of language in interpersonal relations. Sometimes the children offer quite well-constructed sentences, but more often they utter short, primitive sentences or phrases.

AN ECONOMICAL THERAPY MACHINE (ETM)

Our implementation of an automated therapy laboratory on the PDP-10 is clearly not feasible for most clinics, schools or even hospitals around the country. The computer requirements described above would cost well over a million dollars if duplicated faithfully. But this approach is not even desirable. Our goal was to use whatever resources were available to experiment with automated therapy techniques, bypassing as many developmental problems as possible. Now that we have had several years of experience with this type of treatment, we can begin to specify the characteristics of an economical therapy machine (ETM). By
this we mean a machine suitable for use in a large number and variety of institutions dealing with language problems.

First and foremost is effectiveness; the machine must be at least as effective as our system in treating language disorders. Thus it must have many of our system's features which we have found to be important: (1) a video screen and provisions for showing still pictures and (at least crude) motion. (2) An audio device capable of producing continuous sound in synchronization with the pictures. (3) A keyboard or similar device with which the child may select a picture/sound, plus the appropriate logic for selecting his choice from the storage. (4) Storage devices for pictures and sounds; they could use the same device (e.g., movie film with sound track) or separate ones. The storage devices and their accessing logic must have at most a one second delay (½ second is optimal) between the time the child requests a picture/sound and the time it is played for him. The child's attention span will not tolerate a longer delay.

Next in importance is low cost. The machine must be economically feasible for a wide range of institutions; individual machines can and should cost no more than $10000-$50000. It may be possible to group several terminals together under one system, in which case a small computer (e.g., a PDP-8) could be used for the logic. A system with five terminals should cost around $15,000. These are only rough estimates; they might be improved upon.

Other essential characteristics are that the machine be reliable, rugged (the keyboard, at least, must be able to withstand rough handling by children), easily and cheaply repairable (hopefully using stock, readily-available parts), portable, self-contained and requiring no special environmental controls (such as air conditioning or humidity control).

This is admittedly a formidable list of requirements, but one which we believe to be well within the capabilities of current technology. Some suggested implementations are presented below.

There are other characteristics that are desirable but not essential in a therapy machine: (1) Color in the pictures. (2) Better motion than the 2-4 frame motion we use; also more professional cartoon characters. (3) Longer sound than our current three-second time limit. (4) Flashing lights and changing geometric patterns, accompanied by music; the children are fascinated by orderly motion. (5) More physical involvement while participating in the games. Many children are hyperactive and squirm and fidget when required to sit in a chair to play the games. If they could run around the room punching buttons or jumping on pedals to activate the machine, it would not only be more enjoyable to these children but would probably increase their attention span as well. (6) Finally, custom tailoring is important. It should be easy to construct a new game specifically for an individual child. Frequently a child will show interest in a particular sound or word or phrase; it should be possible to construct a new game exploiting this sound for the child's next therapy session.

All of these features are designed to hold the child's interest and to stimulate his involvement with the machine. Variety and novelty are the key characteristics. Any suggestions the machine designer would have would, of course, be welcome.

POSSIBLE IMPLEMENTATIONS OF AN ETM

The following are some rough ideas we have had on the form an economical therapy machine might take.

The video screen could be most simply realized as a standard slide-projector screen. This would exclude some versatility, such as permitting the child to draw on a display screen with a light pen, but the cost would be very low.

The pictures could be generated using a standard film strip driven by a high-speed stepping motor. The motor would have to be fast enough to advance over several frames of the film to get to the desired frame within the ½-1 second time constraints. It should also be able to stop, so that a single frame remains shown, and to proceed at a slow rate suitable for showing motion. Alternatively, a plate of fiche with the pictures imprinted on it would be used together with a scanning head. Both of these approaches require the design of logic circuitry to control and count the stepping of the motor or the positioning of the scanning head. This circuitry constitutes the main non-standard aspect of the ETM. It has to be designed from scratch; but presumably, once the design costs have been met, it could be constructed inexpensively using low-cost integrated circuit chips.

The sound could perhaps most easily be recorded on a sound track on the film strip. This requires that the film strip be constantly moving, even for still pictures. The advantage is that each game could be packaged in a self-contained and easily-changeable cassette. The problem of constructing new games tailored to individual children reduces to the problem of making new cassettes; there is much inexpensive movie-with-sound recording equipment available today. Alternatively, a separate sound tape could be used in conjunction with the picture producing device. This would complicate the controlling circuitry, since two tapes would have to be coordinated (one for the
pictures, one for the sound). A small computer might become feasible with this approach, particularly if several terminals were to be used. Another approach is to have a rapidly rotating drum on which the sound, and possibly the pictures, is recorded. Movable or fixed read/write heads would be used to access the sound; a high RPM would insure that latency is sufficiently small. This arrangement is particularly suited to time-sharing several terminals using a small computer; one copy of the sounds could be used by all the terminals connected to the drum.

We envision a very wide market for a machine designed along these lines. Many hospitals, clinics and schools are becoming increasingly concerned with correcting language disorders. Furthermore, normal children thoroughly enjoy our system. If used in the schools, it could provide a powerful stimulus for their early interest in language, leading perhaps to increased reading efficiency at an early age.

REFERENCES

1 B BETTELHEIM
The empty fortress
The Free Press New York 1967

2 K M COLBY
Computer simulation of change in personal belief systems
Behavioral Science 12 248-253 1967b

3 K M COLBY H ENEA
Heuristic methods for computer understanding of natural language in context-restricted on-line dialogues
Mathematical Biosciences 1 1-25 1967b

4 K M COLBY
Computer-aided language development in nonspeaking children
Archives of General Psychiatry 19 pp 641-651 1968

5 K M COLBY D C SMITH
Dialogues between humans and an artificial belief system
Proceedings of the International Joint Conference on Artificial Intelligence The MITRE Corporation Bedford Massachusetts 1969

6 K M COLBY D C SMITH
Computer as catalyst in the treatment of nonspeaking autistic children
Computer 4 47-53 1971

7 K M COLBY S WEBER F D HILF
Artificial paranoia
Artificial Intelligence 2 1-75 1971

8 PDP-10 system reference manual
Digital Equipment Corporation Maynard Mass 1968

9 O K MOORE
Autotelic responsive environments and exceptional children
Responsive Environments Foundation 1963

10 E M ORNITZ E R RITVO
Perceptual inconstancy in early infantile autism
Archives of General Psychiatry 18 pp 76-98 1968

11 B RIMLAND
Infantile autism
Appleton-Century-Crofts New York 1964

12 D C SMITH
MLISP
Stanford Artificial Intelligence Project Memo AIM-135
Stanford University 1970

13 P SUPPES
The uses of computers in education
Scientific American 215 pp 206-220 1966

14 J K WING
Early childhood autism
Pergamon Press New York 1966