DOCUMENT RESUME

ED 090 498 CS 001 048

AUTHOR	Doublier, Rene M.
TITLE	Interactive Relationships with Computers in Teaching Reading.
PUB DATE	May 74
NOTE	11p.; Paper presented at the Annual Meeting of the International Reading Association (19th, New Orleans, May 1-4, 1974)
EDRS PRICE	MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS	*Computer Assisted Instruction; Elementary Education; *Individualized Instruction; *Reading Instruction; *Reading Programs; Systems Analysis; *Teaching Models
IDENTIFIERS	*Computer Assisted Reading Educational System; Man Machine Communications

ABSTRACT

This study summarizes recent achievements in the expanding development of man/machine communications and reviews current technological hurdles associated with the development of artificial intelligence systems which can generate and recognize human speech patterns. With the development of such systems, one potential application would be the establishment of machine-assisted reading centers, permitting significantly increased individualized reading instruction similar to the techniques employed in modern language laboratories to supplement classroom instruction. The Computer Assisted Reading Educational System (CARES) is proposed as a model for a reading laboratory and described in terms of flow diagrams, system response parameters, input/output displays and devices, estimates of required machine size, system cost, and time development. (Author/RB)

Rene M. Doublier
Univ. of So. California
2915 Hermosa View Drive
Hermosa Beach
California 90254

Interactive Relationships with Computers in Teaching Reading

International Reading Association
Nineteenth Annual Convention
Technology and Reading Session
Friday, May 3, 1974

Interactive Relationships with Computers in Teaching Reading

Rene M. Doublier

I. Introduction

Coincident with the widespread introduction of automatic computational machinery (around twenty years ago), a new science came into being whose fundamental goal was to develop improvements in the techniques of man/machine communication. It soon became obvious to the "computer scientist" that many potential users of digital processors were thwarted by the difficulties associated with mastering the very difficult, low-level, machine-language programming techniques then available, and so (with invaluable assistance from linguists and grammarians), programming system architects set out upon the development of near-natural programming languages which have since become known under such popular mnemonic titles as "Fortran," "Algol," "Cobol," etc. Yet, even with these vastly simplified translators, the ultimate means of communication with machines remained an elusive goal.

> "If computers could only speak their answers as well as display them graphically..."

> "If computers could only hear and understand spoken requests as well as those made on typewriters, punched cards, or magnetic tape..."

The first of these capabilities is commonly referred to as computer-generated or "synthesized" speech, whereas the second (and, incidently, by far the more difficult task) is usually termed automatic speech recognition.

The applications of such a system are too numerous and varied to attempt a complete listing in the present paper. Rather, it is our purpose to acquaint the reader with the fundamental technical problems currently limiting the "state of the art" in man/machine communications, and to present a possible model for a specialized application, namely a "Computer-Assisted Reading Educational System" (CARES).

It is hoped that, in the course of this exposition, some of the problems facing the computer scientist in teaching a machine to speak and to understand human speech will be recognized as being common to the teacher/human pupil situation.

II. Some Fundamental Linguistic Concepts

Speech is a form of communication between human beings which involves the generation and reception of a rather complex acoustic signal. The first major decision facing an architect of a man/machine communication system is that of establishing the fundamental linguistic units which are either to be converted to acoustic energy (synthesized) or extracted from a received waveform (recognized).

The most familiar language units are, of course, words. It might seem natural then to propose that our strategy employ the stored characteristics of a table of words, representative of the base language. However, the awesome size of most unabridged dictionaries of, for example, the English language suggests that inordinately large amounts of computer memory capacity would be required to tabularize the characteristics of words and their various derivatives.

Closer examination of the acoustic properties of human speech reveals that we can express the elemental sounds of a language as a finite set of discrete symbols, commonly called "segmental phonemes."*

Naively then, it would appear that a far simpler and more efficient linguistic representation could be achieved by specifying the language at the phonemic level, and that speech could be

*Approximately 47 phonemes are required for the General American dialect of the English language (roughly described as the /R/-keeping speech of the Mid- and Far West).

produced (or recognized) by a simple concatenation (or segmentation) of these discrete units. Unfortunately, three significant problems associated with the characteristics of human speech severely complicate the segmental phoneme approach.

First of all, continuous speech is not particularly amenable to discrete analysis. The acoustic properties of one phoneme are not isolated and readily segmented from those of the surrounding phonemes. Thus, it is difficult to establish phoneme boundaries, since the characteristics of one phoneme often glide continuously into those of the following phoneme. This, of course, would severely complicate automatic (machine) recognition procedures based solely on phonemic segmentation.

Secondly, these "characteristic" acoustic properties are highly context sensitive. That is, the properties of a given phoneme vary according to the linguistic and acoustic properties of the surrounding phonemes.

Finally, there are significant variations in speech patterns, not only from one speaker to the next, but also in identical phrases spoken by a single speaker on different occasions.

The trade-off is therefore evident. Either we require vast amounts of memory to store our library of larger speech units (words, phrases, sentences, etc.), or we must develop and efficiently program a complex set of context- and speaker-dependent rules for modifying the behavior and recognizing the properties of our smaller (phonemic) elements.* Until computer science develops more advanced and cost-effective techniques of storing and rapidly retrieving the staggering quantity of data required to characterize a diverse vocabulary of words, phrases, or sentences, it would appear that the only feasible technical approach to the synthesis and recognition problems would employ phonemes as basic linguistic units. Once that decision has been made, the next fundamental question to be considered relates to the order of complexity of the phoneme modifier rules.

Research completed at the University of Southern California's Acoustic Phonetics and Hybrid Computation Laboratories has yielded a set of phoneme characteristics and context-modifier rules for General American English which have been compactly programmed in computers of only moderate size. (1) Employing these characteristics and rules, in conjunction with a Terminal Analog Speech Synthesizer, researchers have produced artificially -generated speech of high intelligibility and reasonable natural quality.

While satisfactory solutions to the synthesis problem have been achieved utilizing phonemic-level analysis, the previously mentioned problems of phoneme boundary identification and human speaker variations have severely restricted progress in the area of automatic speech recognition.

*It must also be noted that these rules are language and even dialect sensitive.

III. Some Additional Thoughts on the Recognition Problem

In designing an algorithm for machine recognition of human speech, we might hope to pattern our strategy after our best understanding of how the human brain decodes the speech waveform. Unfortunately, the actual mechanism of human perception and recognition of speech are only rather poorly understood. For example, little is known about the temporal span of the recognition unit. Some researchers argue that the phonemic level is the most likely, while others propose that the recognition process is accomplished at higher linguistic levels, such as syllables, or even possibly words. Since we have concluded that speech is a random process which shows strong evidence of being non-stationary, it is small wonder that past attempts at automatic machine recognition of human speech have met with only very limited success. This is primarily a result of the fact that these experiments have been based, to a large extent, on acoustic pattern recognition, and have made only restricted usage of other available cues and linguistic constraints.

The evidence appears to be mounting that a generalized recognition algorithm (that is, one which is speaker-independent and which assumes no a priori knowledge of spoken text) will be based not only upon the recognition of acoustical patterns of elemental sound units (for example, phonemes), but will also require higher level decoders, such as grammatical and syntactical rules. Although a speech recognition strategy should endeavor to accomplish its tasks at as low a linguistic level as possible, the previously cited arguments would lead one to conclude that, at least on some occasions, it will be necessary to employ the more powerful techniques of analysis at higher linguistic levels. It is the research and development of this minimal set of rules which is the principal stumbling block facing speech researchers.

The foregoing would seem to paint a rather bleak picture for the immediate potential of interactive (man/machine) communications. However, if we now focus our attention on a particular application of this capability, some simplifying assumptions may be made which will significantly reduce the magnitude of the technological problems which must be hurdled.

IV. Computer-Assisted Reading Instruction: The Model

One of the fundamental problems facing our educational system is that of providing adequate individualized reading instruction.

> "Classes containing thirty-five to forty students are, unfortunately, numerous. Teachers of reading in these classrooms complain that they cannot do the job. They mean that they cannot find time for thorough ongoing diagnosis and individual programs for the children who need individual help." (2)

Furthermore, current economic pressures on educational funding are aggravating rather than alleviating this problem.

It would appear then that, just as our universities and colleges have developed systems to permit individualized practice in the large-scale instruction of foreign languages (commonly called "language laboratories"), it is appropriate to consider the application of modern technology to provide a <u>supplement</u> to teacher reading instruction. The following is a brief description of some of the essential and desirable features of a computer-based system for reading instruction.

(1) Pupil Terminals - The pupil terminal includes the interface equipment for displaying written text (either in phonic or lexical form), receiving and converting the pupil's spoken responses into digital data, and the digital-to-analog/synthetic voice tract equipment necessary to produce artificial speech. The graphic display would be best implemented using a cathode ray tube system, with the additional features of a movable cursor to pace the student, an electrostatic pointer to permit the student to synchronize the computer to his position in the text, and a phonic-equivalent display above the lexical form of the word.

(2) Instructional Feedback - The system must have the capability of providing corrective instructions to the student. This, of course, implies that the system must be capable of not only generating high-quality, natural-sounding speech, but also of identifying error patterns or difficulties in pronounciation in the student's spoken text. Note however that this "speech recognition" requirement is a considerably simplified version of that described in earlier sections of this paper, since the computer may be programmed to have *a priori* knowledge of the text. The recognition task is thus reduced to one of comparing the properties of the spoken acoustical waveforms with those internally generated by the computer, using speech synthesis rules. When this comparison exceeds an acceptable

"pronounciation accuracy" threshold, the computer will issue corrective instructions as both graphic and audible corrective responses (phonic, word, and/or phrase).

(3) System Response Time - The system must operate in as near "real time" as possible. Significant delays in input/output processing would prove disastrous in a reading instruction environment. Using as a basis the aforementioned research performed at the University of Southern California (1), it is estimated that a small, special-purpose processor, with 32K main-memory capacity, and one to two microsecond memory cycle time would provide sufficient computational power and speed.

(4) Cost and Development Lead-Time - Excluding initial development costs, individual pupil terminals (including central computer processing) could be delivered (in quantities of fifty pupil terminals) at a cost of around $35,000 per unit. Although this cost may appear excessive, it should be noted that the useful life of the system can be reasonably estimated to exceed five years. The development lead-time of the CARES system, assuming a significant commitment on the part of a major computer manufacturer would be less than five years.

IV. Some Final Thoughts

With every day that passes, our way of life becomes more and more dependent upon computers, whether they be the large general purpose scientific machines that can solve problems in seconds that would occupy several men for years, or the small special purpose devices that perform routine accounting and controlling tasks. And yet, although these machines touch all of our lives in one way or another, their fullest potential to serve mankind will be realized only when we develop methods of teaching computers to communicate in the manner most natural to us, namely spoken language.

References

1. Doublier, R. M., Speech Generation and Recognition Under Hybrid Computer Control, PhD Thesis, University of Southern California (February 1972).

2. Heilman, A. W., Principles and Practices of Teaching Reading, Columbus, Ohio, C. E. Merrill Publishing Co., pp. 16.