

NATURAL LANGUAGE INTERACTION WITH MACHINES:
A PASSING FAD?
OR
THE WAY OF THE FUTURE?

A. Michael Noll
American Telephone and Telegraph Company
Basking Ridge, New Jersey 07920

People communicate primarily by two modes: acoustic -- the spoken word; and visual -- the written word. It is therefore natural that people would expect their communications with machines to likewise use these two modes.

To a considerable extent, speech is probably the most natural of the natural-language modes. Hence, a fascination exists with machines that respond to spoken commands with synthetic speech responses to create a natural-language interactive discourse. However, although vast amounts of research and development effort have been expended in the search for systems that understand human speech and respond with synthetic speech, the goal of the perfect system remains as elusive as ever. Systems for producing natural-sounding speech for large vocabularies with unrestricted grammatical structures and for recognizing spoken speech for large vocabularies with unlimited grammatical structures and any number of talkers are still beyond the state of linguistics and computer science and technology.

Given the problems in the speech domain, it is not surprising that most interactions between people and machines are in the visual mode frequently using alphanumeric keyboards as input and textual display as output. Such visual terminals are already in fairly widespread use in industry and are used for a variety of applications including computer programming, text editing, and data-base access.

The telephone allows speech telecommunications over distance between people. Future visual terminals for the home and businesses will allow textual telecommunications between people. These visual terminals could also be used to telecommunicate with machines in a way that is presently difficult using the telephone and speech.

Viewdata, or videotex, systems are promised soon for the home and will allow data-base access and transactions with machines and textual messages between people. Some viewdata systems use elaborate tree searches to reach the desired frame of information. Some people believe that tree searches will be "unnatural" for many users and some other more-natural language will be needed to search and access these data-base systems.

One conclusion is that the future will see more choices in mode for telecommunications between people and with machines. The choice of which alternate

mode will probably be dependent upon the specific application. For example, textual messages might be both easier to enter by keyboard and to read on a CRT screen than speaking to a recording machine and listening to a recorded message. However, social chatting might be best over the telephone. However, arranging a date with a stranger might be less revealing if done in the textual mode. Considerable opportunities exist for basic research to explore the suitability of these alternate modes for different communications applications.

The fascination of technologists with speech-synthesis chips is about to result in a variety of stand-alone appliances that speak. Ovens that state when the roast is done, washing machines that call for the addition of fabric softeners, automobiles that inform the driver that the door is open, and many other applications will soon abound in the marketplace. In most of these applications, synthetic speech will substitute for a lamp or other form of visual display. The environment will be polluted with the noise of buzzy synthetic speech. Many of these applications will undoubtedly be little more than passing fads.

But in some circumstances synthetic speech will become the way of the future. One example would be synthetic-speech announcements of floors in an elevator thereby eliminating crooked necks!

Most of the preceding examples are very restricted in terms of the language used for the interaction with machines. The problem with unrestricted natural language for communication with machines is that no automatic way has yet been discovered to extract meaning in either the speech or textual mode. The textual mode does eliminate the need for acoustic analysis and hence has been more extensively used in most systems for restricted, specialized applications. However, even if either mode were equally near perfect, questions would still arise about user preference for one mode over the other.

Thus, in the end the future will be decided by the votes of consumers in the marketplace as they choose from the many options presented by technology. The shrewd entrepreneur will use consumer preference and needs to help illuminate in advance the desires and needs of the marketplace. Basic research in linguistics, human behaviour, natural language, and other ancillary fields will have an important role in developing solutions and in understanding people's needs and behaviour.

