

論文 / 著書情報  
Article / Book Information

|                   |                                                                         |
|-------------------|-------------------------------------------------------------------------|
| Title             | Telephone Networks in 10 Years time (Technologies and Services)         |
| Author            | SADAOKI FURUI                                                           |
| Journal/Book name | COST 209 Workshop on Speech Recognition over the Telephone Line,<br>, , |
| Issue date        | 1992, 11                                                                |

## Telephone Networks in 10 Years Time (Technologies and Services)

Sadaoki Furui

NTT Human Interface Laboratories

3-9-11 Midori-cho, Musashino-shi, Tokyo, 180 Japan

### 1. Introduction

One of the ultimate goals of telecommunications is to facilitate smooth communication by overcoming the distance and time difference between people no matter where they are. Although demands for transmission capacity, quality, and speed have almost been met by the recent development of the ISDN (Integrated Services Digital Network), many problems must still be solved if we are to have telecommunication systems that truly overcome the distance and time between people in different places.

Another goal is to create an advanced user-friendly human interface by combining ISDN functions and capabilities with computers. This should be human-oriented interface, with machines stepping up to people, rather than a machine-oriented interface that asks people to step up to the machines.

This paper gives the perspectives of the telecommunication research and development need to reach these goals.

### 2. VI&P Services

NTT (Nippon Telegraph and Telephone) is conducting research and development under the vision of "VI&P" (visual, intelligent, and personal) services for the 21st century. This vision is based on the trend of sociological, economical, and cultural changes and of technological innovation as we move toward the 21st century.

Research and development targets are listed in Table 1. Visual communication services will include high-definition TV that can display natural and clear images, life-sized large-frame display, and multiframe transmission allowing users to choose their favorite frames (multi-motion-picture-windowing). Intelligent communication services will include electronic secretarial services, automatic translation telephony, and various order-made services whose contents can be easily changed by users. These services can be implemented by incorporating advanced information processing capability into the networks. Personal communication services will include chasing communication, in which the network connects calls to each user having personal number wherever he/she is; confidential communication, in which calls can be received only by the specified person; selective receiving that can prevent mischief calls; and automatic charge collection using memory and information processing capabilities of IC cards.

Table 1. Research and development targets for VI&P services.

|                      |                                                     |                                                                                                                                                                                                                                                                                                                                                                      |
|----------------------|-----------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Visual services      | High-definition, large-frame, and multiframe images | Video-shopping, remote observation and medical examination, video databases (electronic museums, fashion information, etc.), high-reality video transmission (life-sized video transmission, etc.), remote environment observation, video notice boards, multisite video conferencing                                                                                |
|                      | Three-dimensional video                             | 3-D video conferencing, high-reality 3-D video transmission, 3-D video relays, 3-D video databases                                                                                                                                                                                                                                                                   |
| Intelligent services | Translation between different languages             | Text-translation communication, translation telephony                                                                                                                                                                                                                                                                                                                |
|                      | Intelligent aid                                     | Voice dialing, time-designated communication, aids for the handicapped (recognition and transformation of characters, speech, finger language, braille, etc.), electronic secretary (responding to telephone calls, schedule management, etc.), database retrieval elicited by ambiguous or natural language reference, notice boards, strategic information systems |
|                      | Flexible networks                                   | Customized services, variable private networks                                                                                                                                                                                                                                                                                                                       |
| Personal services    | Increased usefulness for users                      | Sender name display, calling by person names, automatic collect calls, selective receiving, personal phone numbers                                                                                                                                                                                                                                                   |
|                      | Privacy protection                                  | Confidential communication                                                                                                                                                                                                                                                                                                                                           |
|                      | Flexible charge collection                          | Personal charge collection, personal charge report                                                                                                                                                                                                                                                                                                                   |
|                      | IC-card application                                 | Electronic ticket issue, automatic charge collection                                                                                                                                                                                                                                                                                                                 |

These services will not only efficiently exchange information by simply overcoming distance and time, they will also transmit reality. Communication based on these mechanisms can be performed more precisely and easily. Additionally, various convenient services in our daily life-services such as information retrieval through database management, text editing, games, education, and shopping - will be improved by conversation with computers. These services will be useful not only for businesses but also for creating comfortable living environments and fruitful spare times.

### 3. Human Interface by Multi-Media

In addition to the linguistic information, which is the most important information transmitted in the communication between human beings, such nonlinguistic information as gender, age, physique, action, eye movement, voice characteristics, and facial expressions also play important roles in information transmission. When the communication between people becomes more complex, they tend to use more figures and tables or to point to actual objects. High fidelity (transparent) transmission of this kind of information without distortion or loss is indispensable to the transmission of reality - that is, to making possible information exchange as if both parties were at the same place.

Although most of the information transmitted in the conventional communication (such as telegram, telephone, data transmission, and fax) is linguistic information, future communications will use various media, mainly by combining video and audio media (multimedia or multimodal interface). Tomorrow's interface should efficiently and fully use various media, considering their individual characteristics and keeping the balance of their load in the total system. The human thinking process is expected to be widely supported by the capabilities of an easy-to-use multimedia interface.

Table 2 shows the developing stages of the video and audio interfaces in teleconferencing [1]. A teleconferencing system providing the illusion of reality will become possible as it evolves from the present concentration type to the projection type and then to the three-dimensional type.

If in addition to providing the transmission capability, a new interface system will allow human beings and computers to communicate in the same way that two human beings do, everybody can have a perfect command of computers without special training or learning. At present, information is input from human beings to computers by keyboard, tablet, and mouse, and output from computers to human beings is usually linguistic information provided by displays and printers. Speech and image information have hardly been used. Although human beings are good at understanding words from speech and understanding meanings and contents from figures, computers are so far not

Table 2. Images of evolution in teleconferencing.

|                        | Image | Shape | Concept                                                                          |
|------------------------|-------|-------|----------------------------------------------------------------------------------|
| Concentration type     |       |       | 1-dimensional<br>Users are conscious of equipment                                |
| Projection type        |       |       | 2-dimensional<br>The state of the other party is reproduced on the stage         |
| Three-dimensional type |       |       | 3-dimensional<br>People in different places gather in a virtual conference space |

good at these functions. There is a large gap between the human-computer interface and the human-human interface.

The interface should also be able to accept mistakes by users and to help users whenever they are lost. The future interface should also be able to automatically acquire new knowledge about the thinking process of individual users, to automatically correct user errors, and to predict by accepting rough instruction and inferring details. A hierarchical interface that first uses figures and images (including icons) for expressing global information, and then uses linguistic expression for details will be a good interface matching the human thinking process.

#### 4. Speech Recognition in the Telephone Networks

As described in the previous section, human interfaces using automatic speech recognition technology are expected to play important roles in future telephone networks. Today, many people in the developed countries are employed to simply sit at computer terminals wearing telephone headsets and transfer information from callers into computer systems (databases) - and vice versa (information and transaction services). According to the basic idea that boring and repetitive tasks done by human beings should be taken over by machines, these information-transfer workers should be replaced by speech recognition and synthesis machines. Figure 1 shows a typical structure for task-specific voice control and dialog systems [2].

On the other hand, the un-ergodic arrangement of keys on a standard typewriter was designed to slow typists down in order to prevent the keys of the early mechanical typewriters from jamming. The input to today's supercomputers is thus limited by 19th century technology. Voice typewriting is expected to increase the input speed and to allow many operations to be carried out without the distraction of hand or eye movements away from the focus of the task on the display.

Other applications of speech recognition systems include secretarial assistance, aids for the handicapped (reading aids for the blind and speaking aids for the vocally handicapped), and interpreting (translating) telephony.

If speech recognition systems are to be widely used, they should as much as possible satisfy the following requirements:

- 1) Reach human capability, and
- 2) Exceed human capability (that is, they should be faster, more accurate, more intelligent, more knowledgeable, less expensive, and easier to use.)

As described in the previous section, it is also important to have an optimal division of roles and cooperation in a multimedia environment including images, characters, tactile signals, handwriting, and speech. It is neither necessary nor useful to try to use speech for every kind of input and output.

Future speech recognition technology should therefore satisfy the following requirements:

- a) Fewer restrictions on tasks, vocabularies, speakers, speaking styles, environmental noises, microphones, and telephones,
- b) Robustness against speech variations,
- c) Adaptability to variations due to environmental conditions and speakers,
- d) Automatic knowledge acquisition for phonemes, syllables, words, syntax, semantics, and concepts,
- e) The ability to process discourse (conversational speech) (for example, to analyze context and accept ungrammatical sentences),
- f) Naturalness and ease of use in the human/machine interaction, and
- g) Recognition of emotion.

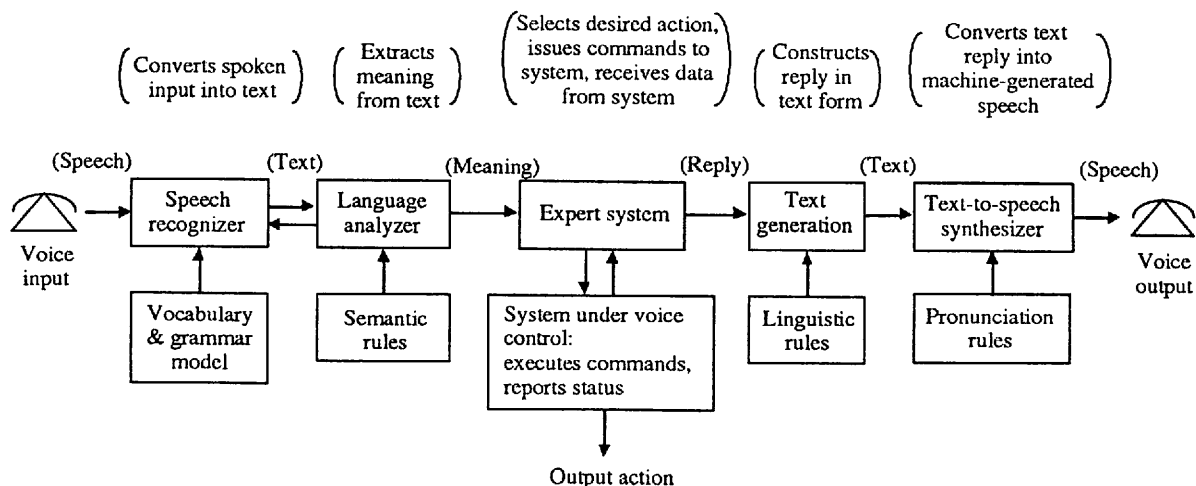


Figure 1. Typical structure of task-specific voice control and dialog systems.

## 5. Discussion

This overview leads us to suggest such discussion topics as

- 1) How can we decide the roles of speech recognition technology in the multimedia environment?
- 2) How to measure the quality of speech recognition systems?
- 3) How much will automatic translation telephony technology be able to contribute to the globalization of international telephone networks within the next 10 years?
- 4) How should the speech recognition and knowledge processing systems be integrated into the telephone networks - centralized or distributed to the user's telephone sets?
- 5) To what extent and how can speech recognition systems be personalized?

## References

- [1] N. Koizumi (1991), "A review of control technology for sound field synthesis", J. Inst. Television Eng. Jpn., Vol. 45, pp. 474-479
- [2] J. L. Flanagan (1991), "Speech technology and computing: A unique partnership", Plenary Lecture, EURO-SPEECH 91, pp. 7-22