Voice and Text Messaging ---A Concept to Integrate the Services of Telephone and Data Networks

Lin-shan Lee and Ming Oun-young

Dept. of Electrical Engineering National Taiwan University
Taipei, Taiwan (Tel) (02)392-2444

Abstract

Although ISDN is one of the major trends for telecommunication development, it is also highly desired to have some intermediate technology which can immediately improve the services of the current networks before actually completing the implementation of ISDN. In this paper, a concept of Voice and Text Messaging System (VTM) is proposed, which can integrate the distinct services of both the telephone and data networks very quickly. In Taiwan, Rep. of China, the telephone network has very wide coverage and large number of users, while the data network has very limited number of subscribers because they have to possess a terminal. The center of VTM described here is a Chinese text-to-speech system which can transform any Chinese text processed in the data network into Mandarin voice, and therefore this voice signal can be transmitted through the telephone network and received by the telephone network users. These users can key in their instructions such as choice of information, text processing, forward and backward skipping by pressing the touch-tone buttons of the telephone set. The electronic mail and database information services provided by the data network therefore become a portion of the voice mail and message sercies provided by the telephone network. The large number of telephone network users, even without a terminal, can thus be served by both networks.

I. Introduction

The information age coming true today can be characterized as a time of exploding demand for communication applications and services of all kinds. These demands create both great opportunities and difficult challenges for the development of advanced telecommunication technologies. One of the major trends for this development is expected to be the Integrated Services Digital Network (ISDN), in which a full range of voice, data, text and video services will be supported. However, the current telecommunication networks were planned several decades ago, and it takes time for these current networks to be eventually changed into ISDN. It is therefore also highly desired to have some intermediate technology which can improve very quickly the services the current networks can offer without actually completing the implementation of ISDN. In this paper, a concept of Voice and Text Messaging System (VTM) is proposed, which can provide all current telephone network users the access to all information services given by the current data network. Although the concept is described in this paper in the form of handling Inform ation in Chinese language, it is definitely not limited to any country or any language.

In Taiwan, Rep. of China, the public switched telephone network has long been developed very well. In average about every three persons share a telephone set, and the network covers almost the entire country, from the mountains to the islands. On the other hand, the public switched data network was established several years ago, but the users are always very limited. Although this data network is supported by very good database information services, only very few subscribers possessing terminals or computers can have access to this network. Because these two networks are completely different, one serving voice and the other data; they can not communicate with each other. The idea of the Voice and Text Messaging System (VTM) is to provide a channel for the large population of the telephone network users to have access to the data network. The center of VTM is a Chinese text-to-speech system which can transform any Chinese text processed in the data network into Mandarin voice, as if read by a human being, and therefore this voice can be transmitted through the telephone network and received by the telephone network users. The users can "key in" the instructions for handling the text by pressing the touch-tone buttons of the telephone set. Operations such as choice of information, message processing, forward and backward skipping, etc. can all be made available. In this way, the large number of users all over the country, even without a terminal, can be served by the data network

0408 CH2538-7/88/0000-0408 $1.00 © 1988 IEEE

through a telephone set. The electronic mail services and voice mail services can thus be combined, and the services of the networks improved. This is considered as an alternative technology which can improve the network services immediately before going to ISDN and without having to make any modification on the current networks. The detailed concepts, approaches, design and operations of this system will be discussed in the following sections.

II. The Basic Concept

The major elements of the current telecommunication networks in Taiwan, Rep. of China [1] are depicted in Fig 1. The public switched telephone network is well developed, having wide spread coverage all over the country and very large number of users. Some computers and terminals can also make use of this network through a modem. The voice mail services are recently provided which is achieved by a voice information center and a voice message service center. The user can receive his voice mail from his voice mail box very conveniently. On the other hand, the public switched data network is relatively new, having only limited number of users because every user has to possess a terminal or computer. Electronic mail services and database information services are provided in this network through the text message service center and the database information center, the latter can provide many usefal public information stored in the database. The above two networks are mutually independent, one serving voice and the other data. They can not communicate with each other. Only very small number of users who are subscribers of both networks and possessing both a telephone set and a terminal or computer can have access to both networks. Therefore most of the large number of users of the telephone network can not be served by the data network.

The concept of the Voice and Text Messaging System (VTM) is shown in Fig. 2, which provides a channel for communication between the above two networks. The center of VTM is a Chinese text-to-speech system [2,3] which can transform arbitrary Chinese text into Mandarin voice as if read by a human being. Such a text-to-sheech system has been successfully developed and implemented, which is the key for the concept of VTM. Although the guality of the synthesized speech is not very satisfactory yet, the intelligibility is very high even after being transmitted through telephone channels. Therefore all text information processed in the data network, including the electronic mail and database information, can be transformed into voice, transmitted through the telephone network, and received by the telephone network users, as long as the text is in Chinese. In this way, we don't have to make any modifications on the current networks, and the voice mail, electronic mail and database information services remain the same. In addition, the electronic mail and database information services can be transformed into voice and become a portion of the voice mail service. For example, the subscribers of the data network can send their electronic mail to a telphone network user who doesn't have a terminal or computer because the mail can be transformed into voice and stored in the receiver's voice mail box. Also, the telephone network users can make use of the database information services by listening to the information stored in the database and read by the text-to-speech system. Therefore the large number of users can all be served by the data network through a telephone set, even without a terminal, and the electronic mail and database information services can be included in the voice mail service. The services of the telecommunication networks can thus be immediately improved.

A nice feature of information in form of text is that the text can be read by the users selectively and repeatedly. The readers can easily skip the parts not interesting and repeat on the parts of special interest to them. This feature should therefore also be implemented on VTM. The operations on the text such as choice of information, message processing and forward and backward skipping can be acomplished. The instructions of the users can be keyed in by pressing the touch-tone buttons of the telephone set. Of course, the convenience and efficiency achieved can not be the same as text processing on the data network, because a user sitting before a terminal can read many words on the screen in a sight simultaneously, but can only listen to the voice word by word. Also, there are many other special features of text information which can't be obtained in VTM. For example, if a telephone user would like to have a copy of the text, he has to go to a different place with a printer available to receive the copy. Furthermore, many types of information such as graphs, figures, tables, paragraphs, size of the characters or letters, punctuation marks, etc. are very difficult to be presented in voice. In other words, by extending the electronic mail and database information services to the entire telephone network, inevitably the level of the services in the telephone network can't be equally high as that in the data network.

III. The System Configuration and Operation Procedures

The simplified block diagram of VTM is shown in Fig. 3. The central part of the system is a Chinese text-to-speech system which will be described in detail later, and a control center which is in charge of the control, management, and operation of the complete system. The telephone network interface receives all messages and instructions from the telephone network and transfer them to the control center, and the data network interface receives all messages and instructions from the data network and transfer them to the text buffer and the control center. The text information to be read by the text-to-speech system will be temporarily stored in the text buffer after obtained by the data network interface from the data network. The voice information synthesized by the text-to-speech system will then be temporarily stored in the voice buffer before transmitted to the telephone network through the telephone network interface. The control center is responsible for the processing of all instructions and messages from both networks and the control of all operations in the system.

When a telephone network user would like to use the database information provided by the data network, he can call VTM and key in the instructions. These instructions will be transferred to and processed by the control center. The control center then notices the data network the desired services via the data network interface. The text information is then provided by the data network, stored in the text buffer, read by the text-to-speech system, and sent to the user through the telephone network interface and the telephone network. The feedback information usually appearring on the screen in the data network such as "The requested information is not found, try again", "For further information, please press 62" can be provided by synthesized voice through the telephone network. The telephone network user can also skip forward or backward a given lenqth of text or choose to repeat a desired text information by pressing previously assigned numbers. On the other hand, when a data network subscriber would like to send electronic mail to a telephone network user, the mail will be sent to VTM. His instructions will be transferred to the control center through the data network interface. The text of the mail will be transformed into voice by the text-to-speech system and transmitted to the voice mail box by the telephone network. The telephone network user will receive the mail just as voice mail, except the voice is synthetic. He can also select, skip or repeat a given part of the mail by pressing previously assigned numbers. Of course, if such a system is to be actually implemented and operated in practical telecommunication networks, special considerations should also be given to the traffic needs, the necessary number of channels, and the implied hardwave design structures.

IV. The Chinese Text-to-speech System

Here we are going to very briefly describe the Chinese text-to-speech system, which is in fact the center of VTM. This system can transform arbitrary given text in chinese into Mandarin voice, and is successfully implemented with satisfactory performance. The design approach is based on a syllable concatenation model due to special considerations on the characteristics of Chinese language [2,3].

There are at least some 13 thousands of commonly used Chinese characters, each character is monosyllabic. There are at least some 60 thousands of commonly used words in Chinese, each composed of from one to several characters. However, the total number of different syllables in Mandarin speech is only about 1300. The use of syllables as the basic units to synthesize Mandarin Chinese therefore becomes a very natural choice. Speech waveforms for Chinese sentences can be synthesized directly by simply concatenating the syllables in the sentences and adjusting the parameters describing the acoustic properties of these syllables. Another very special important feature of Mandarin Chinese language is the existence of the lexical tones. Chinese is a tonal language. There are basically four different tones, i.e., the high-level tone (usually referred to as the first tone), the mid-rising tone (the second tone), the mid-falling-rising tone (the third tone), and the high-falling tone (the fourth tone). It has been shown that the primary difference for the four tones is in the pitch contours, and in fact there exist standard patterns for the pitch contours which will produce the four tones. If the differences among the syllables due to lexical tones are disregarded, only 418 syllables are required to generate all the pronunciations for Mandarin Chinese.

The Chinese text-to-speech system based on the above syllable concatenation concept has a block diagram shown in Fig. 4. In the database the LPC coefficients for the 418 first-tone syllables and the standard patterns for the pitch contours of the four lexical tones are stored. The synthesis rules are a set of general rules which determines how the parameters describing the acoustic properties of the syllables should be adjusted when the syllables are concatenated to form unrestricted sentences with arbitrary text. These rules are the key technology to obtain synthesized voice with

satisfactory guality. They will be summarized very briefly in the following:

(1) The Tone Concatenation Rules

When the syllables with different tones are concatenated in natural speech, the standard patterns for pitch contours are subject to various modifications.For example, if a fourth tone precedes another fourth tone without any pause between them, the first fourth tone will be modified such that the slope of the pitch contour will be decreased by 20%. Also, if a fourth tone is followed by a third tone, the third tone will be modified such that the entire pitch contour should be shifted up to make a continuous contour connecting that of the preceding syllable, etc.

(2) Special Sandhi Rules for the Third Tone

The third tone has a "mid-falling-rising" pattern for the fundamental frequencies. However, such a third tone is produced fully only for special occasions. In many cases only the first half or the second half of the third tone will be produced depending on the syllable it precedes.

(3) Stress Rules and Intonation Patterns

When two syllables form a word, the stress is in general assigned to the second syllable, although exceptions exist in some cases. When more than two syllables form a word, the primary stress is given to the last syllable, the secondary stress to the first syllable, while those in between are least stressed. Also, the intonation pattrn of a declarative sentence is in general declining, etc.

(4) Syllable Duration Rules, Pause lnsertion Rules and Energy Modification Rules

The duration of each syllable should be adjusted according to different factors such as the tone, the initial consonant, the word it forms, etc. Also, pauses of different length should be assigned to different punctuation marks and syntactic boundaries. The energy level of each syllable should be modified based on different considerations such as the tone, stress, etc., too.

The flow chart of the complete text-to-speech system is shown in Fig. 5. The system first extract the parameters for the syllables from the database according to the input text. The syllable duration is then defined, pause inserted, pitch periods adjusted, energy modified, and the speech synthesizer finally produces the output speech. The system is implemented using Digital Signal Processors with the aid of a personal computer. The intelligibility of the synthesized speech is tested and found to be very high, even after transmission through telephone channels. This is why this system can be used to develop VTM.

V. Conclusion

The concept of a Voice and Text Messaging System (VTM) is described in this paper. It can help the telephone network users to have access to the data network services, and make the electronic mail and database information services a portion of the voice mail service. The services of the telecommunication networks can thus be improved immediately.

References

[1] Proceedings of the Telecommunications Laboratory, Telecommunications Laboratory, Directorate General for Telecommunications, Rep. of China.

[2] Ming Ouh-young, Chiu-yu Tseng, Lin-shan Lee, "Design Considerations and Preliminary Results for a Chinese Text-to-Speech System, "1984 International Computer Symposium, Dec. 1984, Tamkang University, Taipei, Taiwan, Rep. of China. pp.1331-1341.

[3] Ming Ouh-young, Chiu-yu Tseng, Lin-shan Lee, "A Chinese Text-to-Speech System Based on A Syllable Concatenation Model, "1986 International Conference on Acoustic, Speech and Signal Processing, Apr. 1986, Tokyo, Japan, pp.2439-2442.

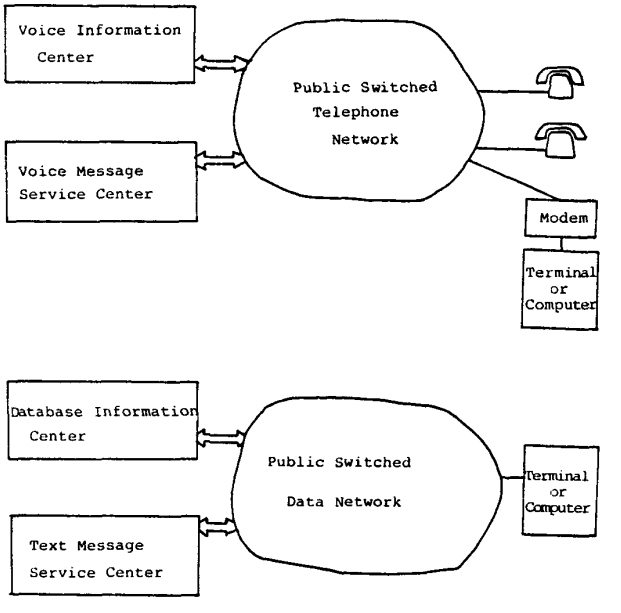

Fig 1. The current telephone network and data network. No communication between the two.

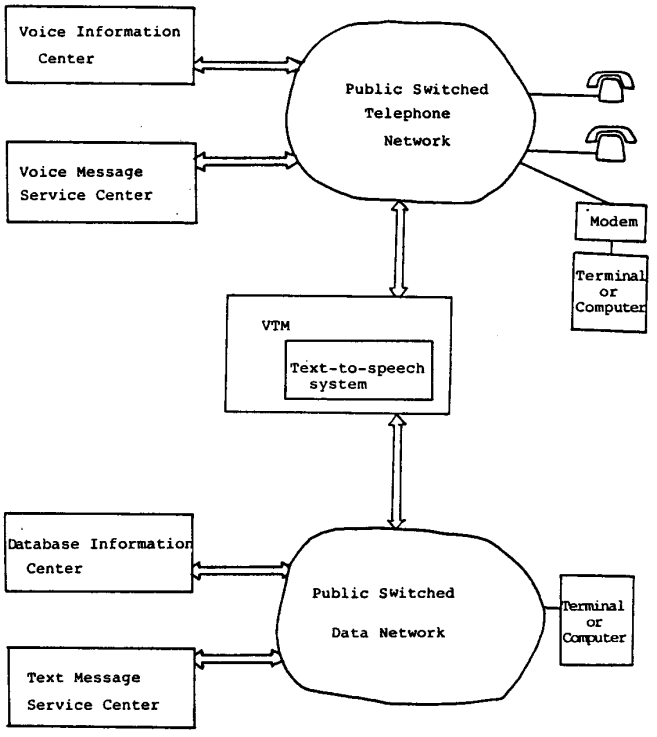

Fig 2. The concept of the Voice and Text Messaging System (VTM)

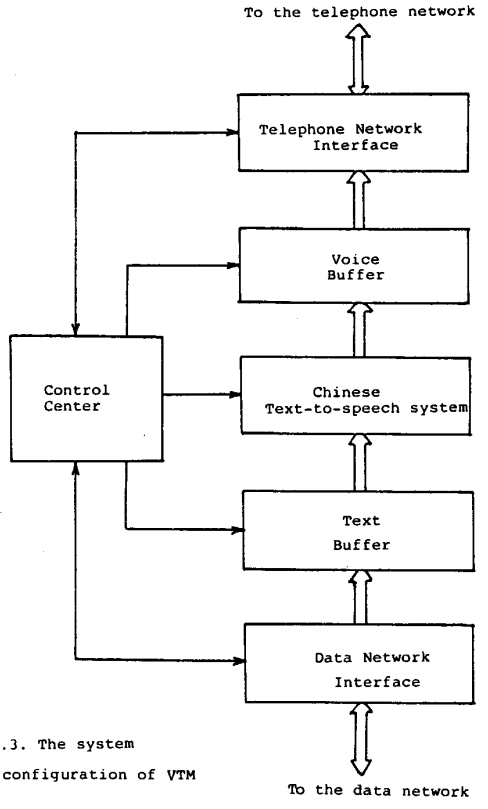

Fig.3. The system configuration of VTM

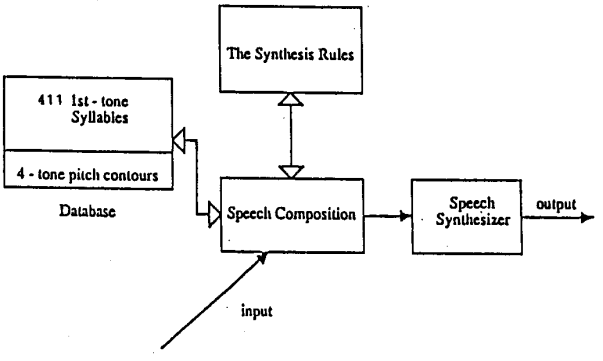

Fig.4. The block diagram of the Chinese text-to-speech system based on the syllable concatenation concept

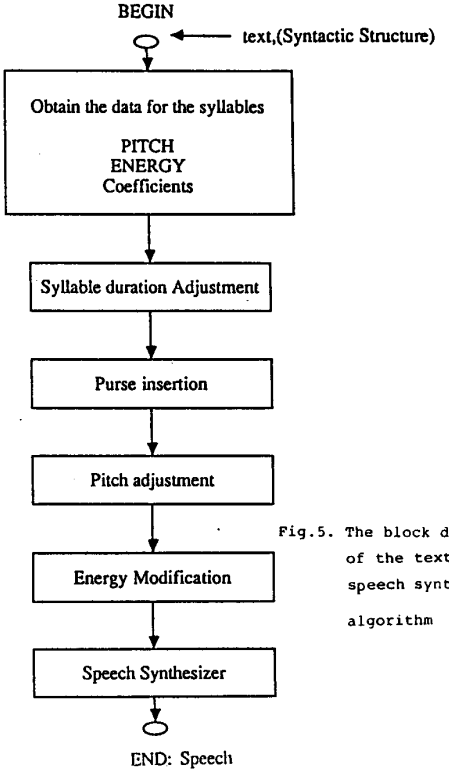

Fig.5. The block diagram of the text-to-speech synthesis algorithm