A Proposed Chinese Language

Videotex and Teletext Standard

Hengsheng Cai[*], J. Spruce Riordon[*], and C. Douglas O'Brien[+]

The three existing videotex and teletext standards are compared briefly. The principal addition necessary for use in China is the capability of supporting at least a subset of Chinese Hanzi characters. This paper proposes a Chinese Presentation Level Protocol Syntax (CPLPS) of videotex/teletext and outlines an experimental videotex/teletext model to implement CPLPS.

1. Introduction

Videotex and Teletext link broadcast means of distribution with computer and display technology to allow the user access through a terminal to textual and pictorial information from data bases. Videotex allows access to very large databases by a relatively small numbers of users. On the other hand, the complementary Teletext technology is used to broadcast relatively small amounts of data to very large numbers of people via broadcast television stations.

Three major presentation level coding standards have achieved international acceptance for videotex. Presentation level coding standards used in four Teletext systems recommended by CCIR more or less agree with those adopted for videotex. In addition there are numerous other private systems, primarily textual, which have been used to distribute information in business, educational and individual applications.

The three internationally adopted coding standards are:

a) NAPLPS, the North American Presentation Level Protocol Syntax applicable to both videotex and teletext. NAPLPS was adopted officially as a Joint ANSI (American National Standards Institute) / CSA(Canadian Standards Association) standard in 1983 [1]. It was also adopted as Data Syntax III of the CCITT Recommendation T.101 in 1984. A fundamental design philosophy of NAPLPS is its independence from a particular terminal equipment implementation technique. An important aspect of this approach is the concept of the unit screen, an abstract rectangular area, or "virtual screen", which maps to the visible display area. Both forward and backward compatibility are thereby assured, and manufacturers are provided with the freedom to produce highly efficient and less costly terminals. NAPLPS is flexible and may be used in various ways depending upon the application.

b) CAPTAIN, the Character and Pattern Telephone Access Information Network from Japan, adopted as Data Syntax I of CCITT Recommendation T.101 in 1984 [2]. The CAPTAIN videotex system was developed primarily to accommodate special Japanese language needs. Its design philosophy is similar to NAPLPS, it makes use of the same unit screen concept. Its PDI set is identical to that of NAPLPS, making conversion and transcoding of information between the NAPLPS and CAPTAIN systems relatively simple.

c) CEPT PLPS, the videotex standard developed by the Conference of European Post and Telecommunications Administrations (CEPT), which includes as components the British Prestel, the French Antiope and the West German Bildschirmtext national videotex systems. It was adopted as Data Syntax II of CCITT Recommendation T.101 in 1984 [3]. The CEPT videotex standard is a composite standard of the various national videotex systems developed or used in Europe. It contains a large array of standardized options, that is, components which may be incorporated into the various national videotex standards.

----
* Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada, K1S 5B6.

+ IDON Corporation, Ottawa, Canada, K1Y 4J8

The videotex and teletext coding standards situation is similar in many ways to that of television broadcasting : three distinct standards exist in a more or less competitive mode. The CEPT photographic scheme is not compatible with that of CAPTAIN, and its geometric scheme is not compatible with that of NAPLPS. A key question with respect to the CEPT standard relates to the definition of the conformance requirements. CEPT is in effect a collection of national standards with some commonality. No country is likely to implement the standard in its entirety, and different countries may implement different subsets. CAPTAIN is a multi-part standard in hierarchical form, and includes Japanese Kanji and Japanese mosaics. NAPLPS is a single unified standard; one cannot meaningfully implement only parts of it. This cohesiveness, together with the support of both private corporations and government agencies throughout the United States and Canada, allows manufacturers to design VLSI chip sets for large volume production.

2. A Proposed Chinese Coding Standard for Videotex/Teletext

A videotex/teletext standard applicable to China is required to meet specific Chinese needs, particularly in the areas of education, transportation and government [4]. Unfortunately current standards cannot be used immediately because of the special needs involved in accommodating the Chinese language. To meet these needs and to allow for international compatability, however, China can build upon the existing presentation level protocols of videotex and teletext, by enhancing them to provide for special Chinese requirements. It is therefore proposed that a Chinese standard be based on the coherent and stable NAPLPS specification and upon the techniques used in the Japanese CAPTAIN system to encode Kanji characters.

The key question for a Chinese coding standard is how add a Chinese Hanzi set capability. In the Japanese CAPTAIN system, a two byte character coding technique is used. However, there are some important differences between the Japanese Kanji set and the Chinese Hanzi set. The major problem is the much larger number of Chinese characters. To be comprehensive, a three byte code table or multiple two byte code tables are required.

The most frequently used Chinese characters number about 3000-6000. It is believed that these characters will be sufficient for teaching, everyday use, newspaper printing and popular catalogs. These Hanzi can be represented by a two byte code table. Other techniques such as geometrically drawn characters or the use of macros or DRCS may be used to accommodate the less frequently used additional Hanzi characters.

2.1 A Chinese Graphic Character Scheme (CGCS)

Several nations have registered two byte character code tables to accommodate ideographic languages. In particular, China has registered such a set of characters to cover the needs of the Chinese language in the People's Republic of China. This set, known as GB 2312-80 [5], contains over 6000 Hanzi characters, more than twice as many as the CAPTAIN Japanese-Kanji code table.

The first half of the Chinese Hanzi graphic character set (CGCS) includes the 3755 most frequently used Hanzi characters ordered according to the Chinese phonetic symbols. The second half comprises the 3008 next most frequently used Hanzi, ordered according to the established radical and stroke count sequence, subsorted by stroke order.

2.2 Chinese Presentation Level Protocol Syntax (CPLPS)

The presentation level coding standard which is proposed for use in China consists of graphic character sets extracted from the CCITT standard on Videotex T.101, primarily data syntax III (NAPLPS) along with the Chinese two byte character code table for Hanzi from GB2312-80. CPLPS should include the following sets :

a) Chinese Graphic Character Set: 2-byte Hanzi code table (GB2312-80);
b) primary character set (international reference version derivative including the YEN sign as the monetary symbol or GB1988);
c) supplementary set from T.101 Data Syntax III;
d) PDI set from T.101 Data Syntax III;
e) macro specification from T.101 Data Syntax III;
f) DRCS from T.101 Data Syntax III.

The proposed CPLPS displays a high degree of compatibility with the standards of North America and Japan. In particular, it makes use of a scheme similar to that of CAPTAIN for the production of Kanji characters, and it implements as its graphics set the PDI setof NAPLPS. The

latter, sometimes referred to as 'the ASCII of graphics', shows promise of widespread use in education and business. The principal disadvantage of the proposed standard is its incompatibility with CEPT, a factor which must be weighed against the benefits of access to the information services available in North America and Japan.

3. An experimental Videotex/Teletext model

Fig. 1 shows the principal components of a videotex/teletext system. Its most distinctive component is the information provider system (IPS), a text and graphics system which allows users to create, edit and display videotex frames. Central processing and database facilities are provided by a videotex/teletext centre, and user terminals are linked to the centre through a suitable communications medium.

The authors are at present testing a prototype IPS configuration which can accept as input a variety of character types, including Hanzi, Roman and Cyrillic. Text and graphic input can be mixed. Graphics page creation is handled by means of Createx-Plus, a product of the Canadian network TV Ontario. Input characters are created with KIM, an intelligent Hanzi input method developed in conjunction with Kaihin Research Inc. of Toronto. The system runs on an IBM PC XT/AT under MS DOS. An alternative operating system for fuller compatibilty in a Chinese environment is CC DOS, an MS DOS compatible system developed in China.

Users of KIM enter Hanzi characters as a set of radical strings. By a combination of forward and backward chaining, frequency-of-use statistics are used to generate the most likely character from a partial radical sequence; the character then appears on the screen. This built-in intelligence allows the system to accept out-of-sequence radical strings, and to tolerate to some extent input errors such as missing and/or superfluous strokes. The average number of keystrokes per character has thus been reduced to approximately 2.9 - 3.1. In addition, it is not necessary to memorize an input coding technique. Trials to date indicate that this method is easy to learn, requires relatively little effort, and leads to fairly rapid input of Hanzi characters through a standard keyboard [6].

4. Conclusion

Videotex/teletext services and applications are spreading throughout North America, Western Europe and Asia. The technology appears to have considerable application in China. The special requirements of the written Chinese language, however, necessitate an augmentation of existing standards to accommodate the Hanzi character set. This paper outlines a Chinese Presentation Level Protocol Syntax capable of supporting over 6000 Hanzi characters. A prototype videotex system is also described which makes use of recently developed software to implement a version of CPLPS. It can be concluded that, although further development is necessary, CPLPS technology is fully practicable.

REFERENCES

1. ANSI S3.110/CSA T500, "Videotext Teletext Presentation Level Protocol Syntax (NAPLPS)", 1983.

2. CCITT Recommendation T.101, "International Interworking for Videotex Services, Annex D", Malaga-Torremolinos, 1984.

3. ibid., Annex C

4. H. Cai, C.D. O'Brien, J.S. Riordon, "A Proposed Scheme for a Chinese Videotex Standard", Computer Processing of Chinese and Oriental Languages, vol. 2, no. 4, 181-197, October 1986.

5. Code Table of Graphic Characters GB 2312-80, China State Bureau of Standards.

6. A. Chiu, F. Wong, H. Cai, "An Intelligent Knowledge-Based Chinese Input System", International Conference on Chinese Information Processing, Beijing, August 1987.
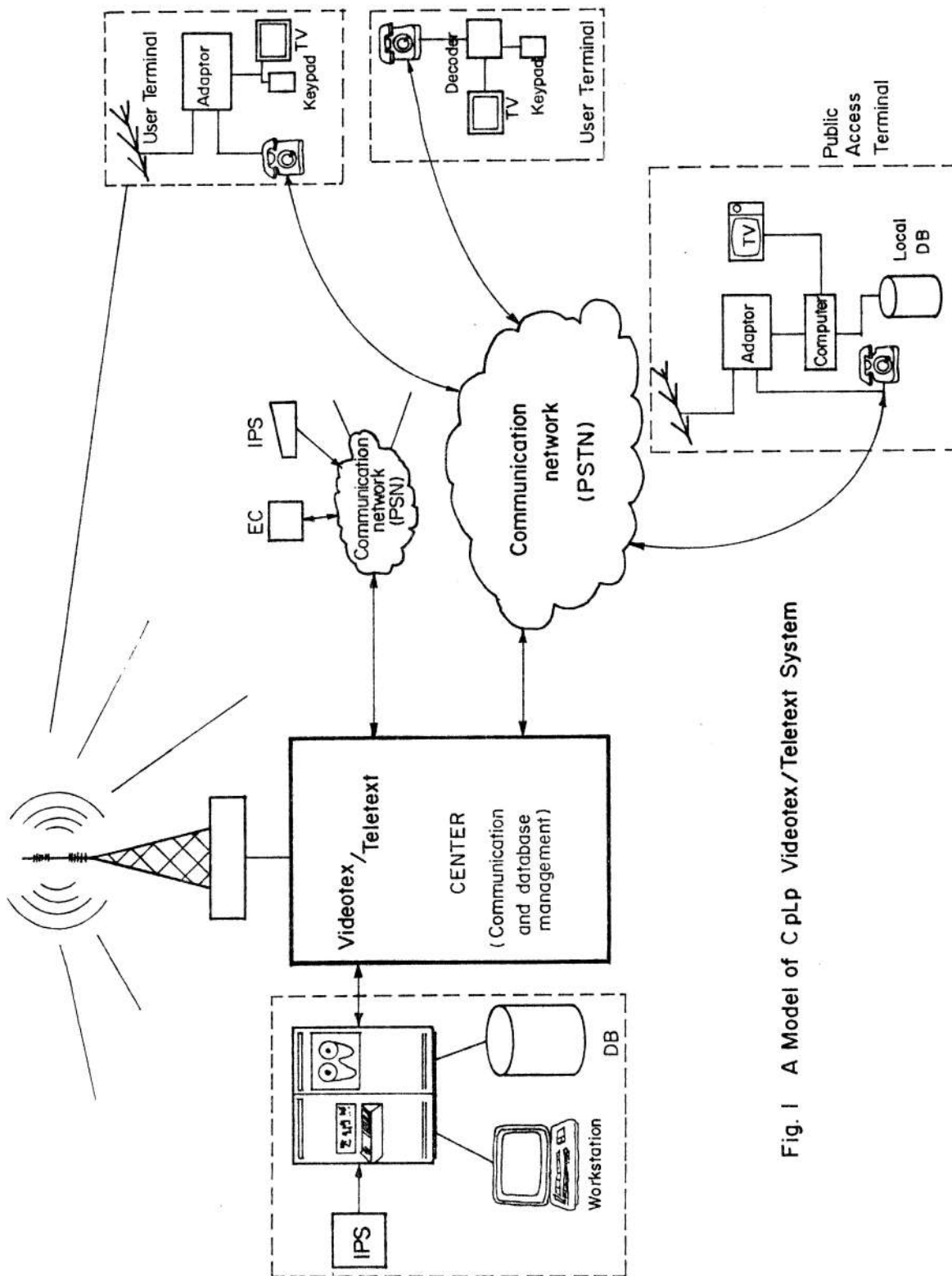
Fig. I A Model of CpLp Videotex/Teletext System

User Terminal
Adaptor
TV
Keypad

User Terminal
Decoder
TV
Keypad

Public Access Terminal
TV
Adaptor
Computer
Local DB

IPS
EC
Communication network (PSN)

Communication network (PSTN)

Videotex/Teletext
CENTER
(Communication and database management)

IPS
Workstation
DB