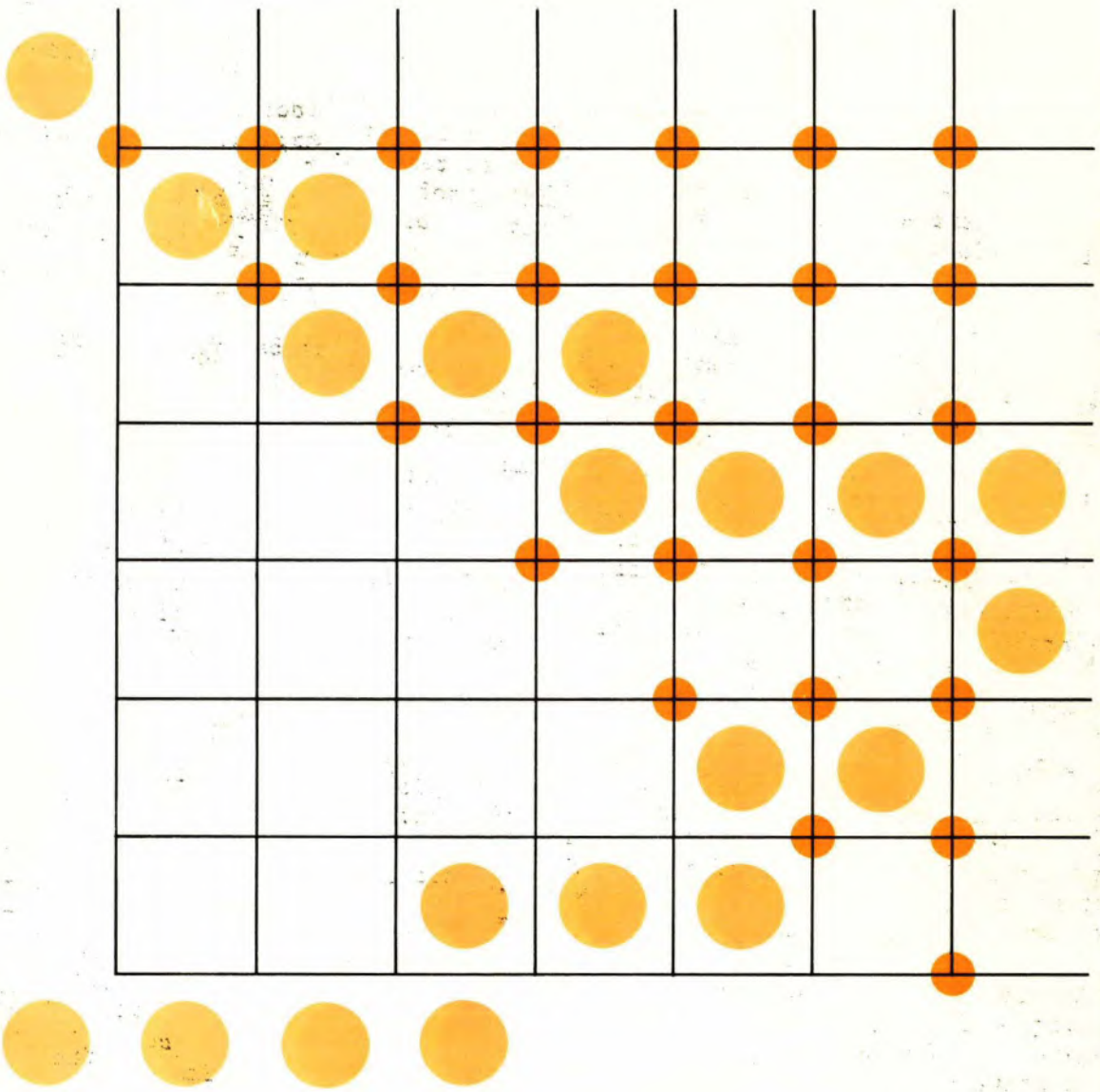
VOICE INPUT/OUTPUT INCORPORATION INTO THE EXISTING TELIDON/VIDEOTEX SYSTEM CONCEPT

Principal Investigator: Dr. G.O. Martens

Research Associate: Victor Shkawrytko



Department of Electrical Engineering
The University of Manitoba
Winnipeg, Manitoba, Canada
R3T 2N2

3

VOICE INPUT/OUTPUT INCORPORATION INTO THE EXISTING TELIDON/VIDEOTEX SYSTEM CONCEPT

Final Report

for

DEPARTMENT OF COMMUNICATIONS
Ottawa, Ontario

DSS Contract No. OSU82-00217

July 1, 1982 - June 30, 1983

Scientific Authority:

Dr. Mike Sablatash
Communications Research Centre
Department of Communications
Ottawa, Ontario

Ottawa, 1982.

Principal Investigator:

by Dr. G. O. Martens
Department of Electrical Engineering
University of Manitoba
Winnipeg, Manitoba

and Research Associate:

Victor Shkawrytko
Department of Electrical Engineering
University of Manitoba
Winnipeg, Manitoba

TABLE OF CONTENTS

ABSTRACT

The incorporation of Voice Input/Output (Voice I/O) into Telidon is studied. A structured approach to system definition is employed using the ISO OSI reference model. Functional requirements are derived using this approach and a terminal subsystem with voice capabilities is presented to illustrate the implementation requirements. Several wideband voice encoding techniques are suggested for the terminal subsystem to satisfy voice quality requirements. Voice recognition implementation issues are also discussed.

Voice I/O is viewed as a member of a set of Integrated Services in order to allow inter-service compatibility and more cost-effective use of the system voice capability. The effects of different distribution media (such as the digital telephone network or Cable TV) on Voice I/O incorporation are also addressed.

ACKNOWLEDGEMENT

The authors thank Dr. M. Sablatash of the Communications Research Centre for suggesting the topic of this research and for pointing-out many of the references referred to in the report.

I. INTRODUCTION

The Telidon System is a public access information system which provides subscribers with the capability of accessing and displaying computer stored data. Access to this system is via common carrier communication lines or other interactive networks.

Current Telidon specifications do not include provision for the use of voice response and/or voice recognition in the exchange of information between user and host computer. Voice Input/Output (Voice I/O) as an interface to a computer is in many cases a desirable feature if it simplifies the use of the system through a more natural form of communication.

A large variety of technologies exist for the analysis and processing of voice signals. These technologies vary in hardware complexity, computational requirements, cost and subjective performance. The incorporation of Voice I/O technology into the Telidon System concept carries not only the problem of selecting a technology for implementation, but involves an analysis of the overall needs of users using Telidon and other services which will become available through a home subscriber terminal. As well, factors affecting the acceptance of a given service from both social and economic viewpoints must be considered in the decision-making process regarding the incorporation of a new service.

The purpose of this study is to survey and resolve the details of incorporating Voice I/O into the existing Telidon Videotex system concept. The methodology used to initially specify the required functions of Voice I/O follows the network layering approach. Specifically, a top-down design approach using the ISO OSI reference model is used to derive implementation requirements. Some protocol specifications are given in the appendix. Initially, a description of Telidon applications is given and

general implementation requirements are obtained. The layering approach gives an indication of the type of protocol issues which need to be dealt with in the incorporation of Voice I/O.

Other than from the network point of view certain practical issues are addressed regarding terminal design. Encoding techniques for speech are surveyed with respect to coding schemes which suit the requirements of the system. Feasible approaches are obtained and a discussion of their implementation within a terminal subsystem is given.

Voice recognition is discussed from a general point of view and suggestions as to the placement of a recognizer within the system are made. The lack of inexpensive recognition equipment at present does not warrant incorporating voice recognition at the terminal level.

In considering the use of voice in Telidon the need for standardization of protocols and techniques is recognized. In order to most efficiently implement a voice service in Telidon, other services must be considered. A discussion of alternate distribution facilities for integrated home services is included to show that Telidon is likely not to develop in isolation and that the selection of a technology for Voice I/O should include considerations for compatibility with other services such as digital telephony or digital television.

The results of this study indicate that incorporation of Voice I/O into Telidon is a multi-faceted problem with many possible directions which may be taken. The fundamental requirement of such an endeavour is the early selection and definition of standards and protocols without the exclusion of other services' requirements.

II. SYSTEM DESIGN METHODOLOGY

In approaching the incorporation of Voice I/O into Telidon, some systematic procedures must be employed to specify the requirements and possible implementations for such a system. The incorporation of Voice I/O from a layered network architecture point of view yields such a systematic procedure. Specifically, layering a Voice I/O system using the ISO OSI Reference Model generates an adaptable and implementation-independent design.

Either top-down or bottom-up design approaches to a network architecture may be employed, both yielding possible problems with unfulfilled user requirements (bottom-up) or unrealizable communication mechanisms due to technological limitations (top-down) [1]. As a result, a reiteration of both approaches must take place several times in a design in order that a technologically realizable system matched to user requirements is obtained.

In this study, an initial top-down approach is employed. Since the technology for current Telidon systems exists and is in commercial use [2] (other than Voice I/O systems) the risk of a technologically unrealizable top-down design is greatly reduced. On the other hand, since much of the technology associated with Voice I/O systems is recent and relatively expensive, practical systems from an economic point of view are a major limiting factor in realizing a system with Voice I/O.

It must be taken into account that a layered network design is primarily an inter-system communication problem with consideration given to user communication requirements. Services to the application layer are provided by the network in order to support the application(s) running within that layer. Communication requirements for Voice I/O are dependent on the type of service that Voice I/O provides to the user. For example,

voice associated with picture data strictly for synthesis purposes (one way communication) and implemented using a complex, low data-rate encoding technique would require different communication services than a two-way digital telephone network. It is therefore necessary to assess the application environment of Voice I/O for Telidon through the study of Telidon application groups in order for clear communication performance requirements to be derived. Given these performance requirements, a layered network design, top-down approach may be pursued.

The following section provides a description of Telidon application groups. The results of this section are taken from [3].

It will become clear with the definition of the applications groups that Telidon is capable of supporting (or shall become an integral portion of) an Integrated Home Services network or services used in an Automated Office. It is for this reason that a further qualification of the system technology base is made in this design. If Telidon Voice I/O is layered into an Integrated Services Digital Network (ISDN) Architecture then a greater degree of flexibility is attained with respect to implementation. Possible future needs are better accommodated with this approach, since system requirements will include those needed for support of an integrated set of services (rather than Telidon in isolation).

Currently ISDNs are in the trial stages of implementation [12,13]. In addition, several somewhat ad hoc approaches to Telidon services distribution exist, although functional service network models have been derived [2]. This should not necessarily limit the design of a layered Voice I/O system within the OSI model. On the other hand, it is useful to relate the layered design to existing in-use network configurations, thus possibly allowing early incorporation of the layered approach into existing networks.

The following section contains a brief description of the ISO OSI reference model. Following this is a summary of the design methodology which is used for layering a network using the OSI model. Following these sections, the implementation of the design is presented.

II.1 ISO REFERENCE MODEL FOR OPEN SYSTEM INTERCONNECTION (ISO OSI)

The International Organization for Standardization (ISO) has developed a model for open system interconnection with the objective of standardizing the rules of interaction between open systems. The standardization of rules is with regard to the external behavior of a system as seen by other interconnected systems [4]. This section begins with a brief description of network layering as it relates to the OSI reference model. Following this is an overview of the OSI model layers.

II.1.1 Network Layering

In order to simplify the standardization of system interconnection, a layering technique was applied to the definition of a network structure. Layering provides a network structure composed of a succession of layers with each layer isolating the upper layers from lower layers. Each layer is attributed a set of functions which complement the functions of lower layers. The layered model is shown in Figure 1.

Each network, if designed accordingly, may be conceptually segmented into layers as shown in Figure 1. Thus, two such networks which are to be interconnected contain corresponding layers. Each layer is said to contain collectively all the subsystems of the rank of that layer within the network. Correspondingly, each subsystem of the layer contains one or more entities.

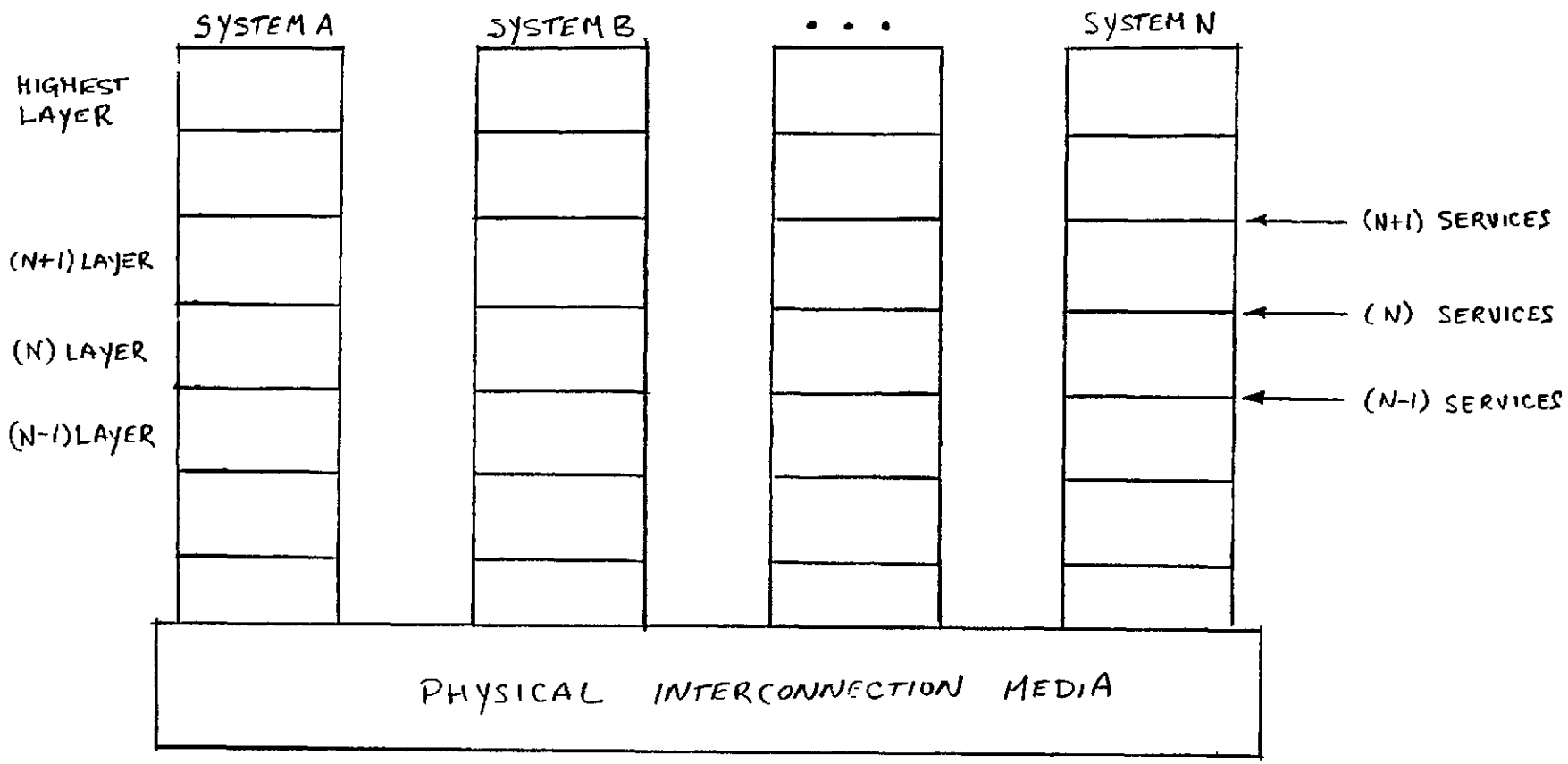


FIGURE 1 - LAYERED MODEL FOR SYSTEM ARCHITECTURES

Entities within a layer are called peer entities and represent the distributed processing capability of the layer to perform its designated functions. The interface between layers may be defined as the set of services that a lower layer provides to the next higher layer. These services are defined such that the specific implementation of a given service is not directly related to the service. In this way, entire network layers may be reimplemented in dramatically different ways with no detrimental effects to the network (providing the same set of interlayer services is provided). How entities within a layer communicate and cooperate to provide the services of that layer is determined by that layer's protocol.

The basic structure and terminology for a layered network model has been presented. Now, a brief description of the OSI model terminology and layer services is given.

II.1.2 THE ISO OSI MODEL

The ISO OSI model architecture is shown in Figure 2. The OSI architecture is made up of seven layers. These are described as follows:

1. The Application Layer: Layer 7
   - highest layer in the OSI architecture
   - all other layers exist in order to support the functions of this layer
   - functions of this layer are to initiate, maintain, terminate and record data concerning the establishment of connections for data transfer among application processes
   - an application is composed of application processes which intercommunicate using application layer protocols. The execution of an application protocol is an application entity. The remainder of the application process is beyond the scope of the layered model.

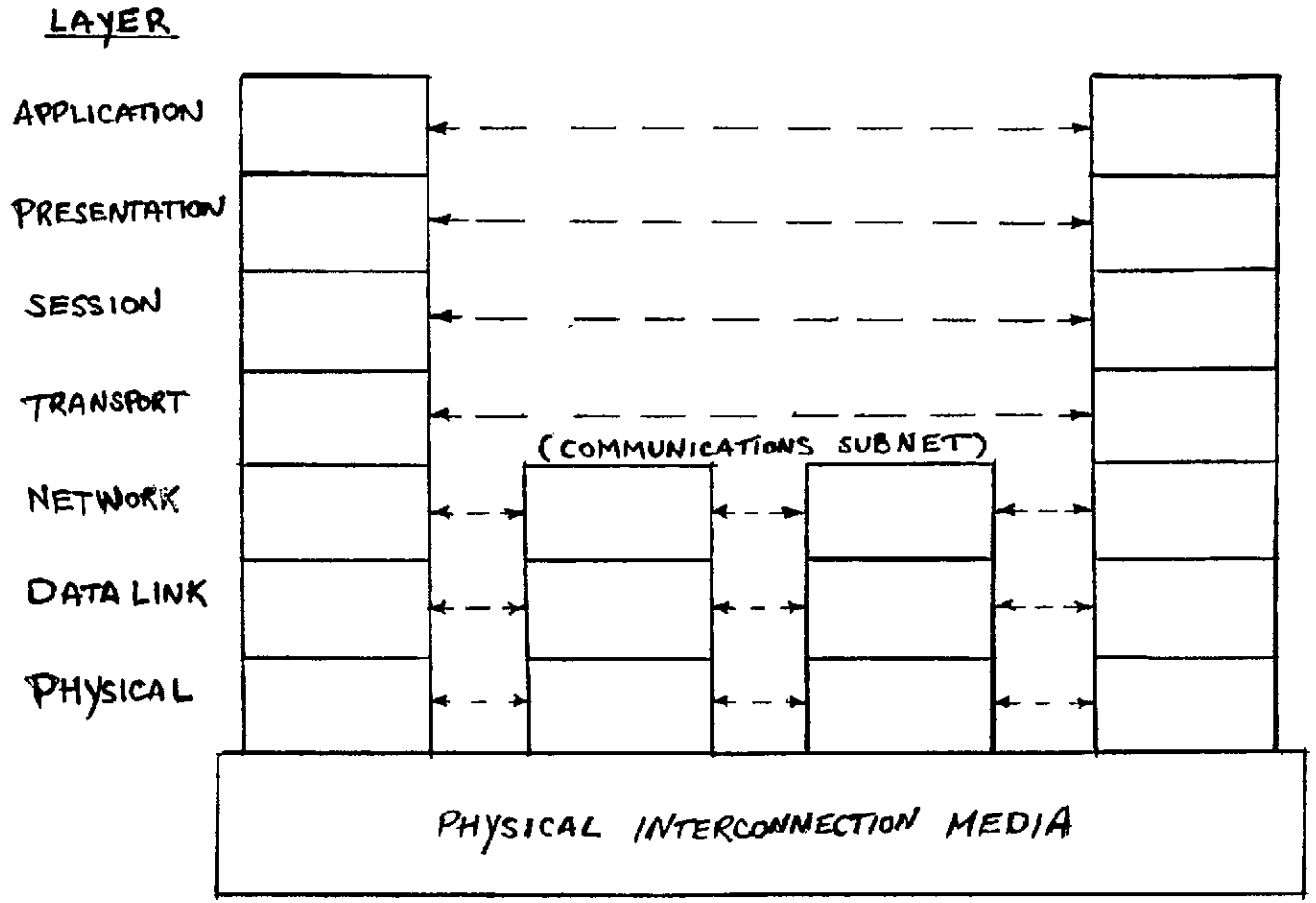


FIGURE 2 - THE ISO OSI REFERENCE MODEL ARCHITECTURE

2. The Presentation Layer: Layer 6

- supports a set of services used by the Application Layer
- these services are for the management of entry, exchange, display and control of structured data
- the Presentation Layer service is considered location independent. Presentation entities communicate via the Session Layer (Layer 5)
- the Presentation Layer provides communication services to applications with minimal interface variability, transformation or application modification.

3. The Session Layer: Layer 5

- assists in supporting interactions between cooperating presentation entities
- services at the Session Layer are classified into the following two categories:
  i. Session Administration:
     binding and unbinding a relationship between two presentation entities.
  ii. Session Dialog Service:
     control of data exchange, delimiting and data operation synchronization between two presentation entities.

4. The Transport Layer: Layer 4

- supports a universal data transport service in association with the services of lower layers
- provides transparent data transfer between session entities
- performs the optimization of use of available communication services for maximum performance/cost ratio for communication between session entities.

5. The Network Layer: Layer 3

- supports routing and switching services to the Transport Layer, thus providing functional and procedural means to exchange network service data units (e.g. packets) between transport entities over a network connection.

6. Data Link Layer: Layer 2

   - supports a functional and procedural means to establish, maintain and release data links between network entities.

7. The Physical Layer: Layer 1

   - provides the mechanical, electrical functional and procedural characteristics to establish, maintain and release physical connections between data link entities.

II.2 METHODOLOGY FOR ISO OSI LAYERED SYSTEM DESIGN

The following approach may be taken to specify a design of an ISO OSI modelled system [5,1]:

1. List user requirements:
   - consider required functions, performance, availability and flexibility.

2. Decompose these functional requirements into layers:
   - use criteria such as physical boundaries and available technology for this decomposition.

3. Design specific algorithms for each function at each level:
   - consider the goals and needs of user and system.

4. Design protocol control information for each level to carry out the algorithm:
   - results in the Protocol and Services units.

5. Examine layer-to-layer interaction:
   - investigate available technologies and services.

6. Reiterate the process until a balance between user interface, system goals and available service mechanisms are reached.

For a given familiar system, a simplified approach may be taken where an initial list of layer services and functions is made (as much as possible) with a thorough examination of functions and services provided by competing systems and possible new features. From this point, a top-down design approach is taken where protocol control information is generated from the application level down. This information is what is necessary for each function and service at a given level to exist. This process results in a complete protocol data unit at the physical layer.

II.3 TELIDON APPLICATIONS [3]

User services provided by videotex systems may be classified according to their primary character of use; i.e., information retrieval vs transactional. Also, accessibility of the services (privately owned vs publicly owned) and target markets for the information (residential vs business) serve to classify videotex system applications. Unfortunately, not all applications fall within these categories. A summary of existing or realistically proposed application areas is given below in order to facilitate the definition of a set of user requirements for layering.

Application Areas:

1. Information Retrieval:

   Examples: catalogs, library references, news, entertainment programs.

   Attributes:
   - page-oriented data bases
   - expressly for videotex
   - dedicated computer
   - heavy downstream traffic

   User Voice Requirements:
   - since retrieval of information is the single purpose here, voice may be used in accessing the information (voice recognition) or voice may be imbedded within the information itself (voice synthesis).
   - access via voice implies some sort of user interface independent of the voice information which may be associated with the retrieval information. Hence, a voice user interface function is required.
   - downstream traffic is heavy if user must access many pages to find desired information. Voice interfacing may decrease the need for irrelevant pages to be sent to the user through an 'intelligent' access algorithm.

- user diversity and vocabulary requirements are characteristic of a public access interface. The specific requirements are discussed later in the study.
- for voice data imbedded within the picture information (retrieved from the data base) a protocol must be defined for decoding - probably a different format than that used in the user voice interface. This is a terminal voice protocol.

2. Commercial Transactions:

Examples: teleshopping, home banking and bill paying, ticket reservations.

Attributes: - applications typically run on service provider's computer connected to videotex network via a gateway.
- downstream and upstream traffic more evenly distributed.

User Voice Requirements:

- transactions must be accurate where the identity of the user is known. This may imply speaker recognition, but at least requires a great degree of accuracy in the transaction algorithm associated with the speech recognition systems.
- a user voice interface may be employed since voice access could provide transaction information with little need for picture information.
- if the transaction algorithm is implemented on the service provider's computer, this must be structured in a compatible format to allow the voice interface to function; i.e., the information transfer must suit both the service provider protocols and the videotex voice interface protocols. This would be handled partially by the gateway and would be partially designed into the voice interface.

- user diversity and vocabulary requirements are of a public access interface.
- data base voice should follow the terminal voice protocol.

3. Advertising and Interest Matching:

Examples: jobs, real estate, marriage broker.

Attributes: - may be as simple as data retrieval or may allow a user response page
- may employ a search between "wanted" and "available" data bases where relational data bases may be used.

User Voice Requirements:

- if information retrieval, then the user voice interface is used. Depending on the sophistication of the user voice interface, much of of the transaction may involve voice, especially if personal information input is required; i.e., prompts for name, address, etc. may be in voice. Otherwise, the voice interface may simply route the user to the desired advertisement or service.
- other than voice interface, voice may be associated with picture data as in application area #1.
- vocabulary and user diversity are of a public access interface.
- voice associated with the retrieved information is encoded in a form compatible with the terminal voice protocol.

4. Messaging and Electronic Mail:

Attributes: - would most likely run on a service computer or may be connected to the network via gateways
- storage requirement for mailboxes
- text editing facility required
- terminals equipped with keyboards.

User Voice Requirements:

- this approach implies text messaging, although voice store-and-forward messaging may be implemented as an alternative. (This may not be entirely desirable since hard copy may be required.)
- if voice messaging is desirable, an encoding technique must be employed for voice digitization, transmission, (storage) and reconstruction. The encoding scheme may be (and most likely will be) common to the user voice interface hardware. Both bit rate and coder complexity are traded off in this case to minimize cost.
  Note: such hardware may be a candidate for digital telephony on the same network.

5. Teleconferencing:

Attributes: - real-time communication between a group of subscribers
- may be video, audio, telewriting or typewritten communication
- prime considerations are channel bandwidth and terminal requirements.

User Voice Requirements:

- real time communication may require a fairly sophisticated communication facility if the phone network is not being used, since arbitrary user nodes must be interconnected
- voice encoding and transmission may be performed similarly to a digital telephone network
- bit rate and communication facility are prime considerations.

6. Education and Computer Aided Instruction:

Attributes: - provision must be made for frequent interaction and enhanced graphics capabilities.

User Voice Requirements:

- highly adaptive voice recognition system if used for interactive learning due to a larger number of untrained users and wide variety of subjects

- an extension of the user voice interface may be employed to implement a set of interactive learning tools
- would most likely need a dedicated machine per user session for complete voice interface due to the number of users and varying subject material
- large data base must be constructed especially for this task.

7. Access to General Computing Facilities and Computer Networks:

   Attributes: - allows access to the power and capabilities of large computers and data banks.

   User Voice Requirements:

   - information transfer must conform to terminal voice protocol if voice is to be used.

8. Telesoftware:

   Attributes: - makes use of the microprocessor in the videotex terminal such that software may be executed in the terminal. The software is downloaded from the service computer. This allows "off-line" (with respect to the network) execution of user applications, thus freeing the application computer and telecommunication circuits.

   User Voice Requirements:

   - software may use the voice hardware on the terminal as an added feature to some downloaded software
   - terminal voice protocol must be known.

In examining the user voice requirements the following voice capabilities satisfy the application area voice requirements:

(a) voice I/O shall act as the interactive user interface to the service (software requirement);

(b) the hardware (and possibly some of the software) available from the user interface in (a) may be used to facilitate other services such as messaging, teleconferencing and digital telephony (hardware requirement);

(c) a voice terminal protocol must be devised to support different voice synthesis (or coding) schemes depending on the use; i.e., videotex picture data plus voice vs digital telephony as an example (software and hardware requirement).

From these three requirements the following functional capabilities of the Voice I/O function must be achieved:

i. Voice digitization, storage, reconstruction and processing capabilities must be available in the network. Such processing may be distributed or centralized depending on both the logical and physical network topologies, and should be capable of near real-time operation.

ii. The hardware for digitization and reconstruction of speech should be reconfigurable for multiple coding schemes or code conversion capabilities should be supported by the network protocols for inter-service speech operations.

iii. Algorithms for operating the User Voice Interface may operate as an independent application process yet accommodate other application processes through appropriate protocol elements.

II.4 DECOMPOSITION OF FUNCTIONS INTO LAYERS

The next step in the design methodology of the OSI model is the decomposition of functions into layers. Here, several criteria may be used for this decomposition [1]:

(a) communication functions are divided according to physical and logical boundaries within the system;

(b) functions are divided based on the change of addressing or multiplex information into a communication mechanism, e.g., channels are addressed individually on a multiple-access channel; individual nodes are addressed in a distributed network; individual users or processes are addressed within a node. These individual addressing changes may be used as layer boundaries;

(c) functions are divided according to their commonality; and

(d) functions are divided according to what services and interfaces are already available (or will be available).

The decomposition approach taken in this study follows the simplified design method mentioned earlier along with criteria (a), and (d) mentioned above. Specifically, a list of layer services is made with the Voice I/O functions in mind based on physical and logical system boundaries and available interfaces and services.

II.4.1 The Application Layer

The Application Layer should contain all operations associated with the User Voice Interface from an application process point of view.

This layer deals with the user-system dialog; i.e., the "syntax" of the User Voice Interface. This syntax must be constructed with both technological and behavioral aspects in mind. The boundary between this layer and the lower layer is drawn at a "virtual voice command channel". Specifically, the lower layers provide a service of digitizing voice spoken into a microphone, recognizing it, translating it into a command and responding appropriately with voice from a speaker. The application level functions assume this service. The application entities use these translated commands in the operation of the User Voice Interface.

The application level functions are those of a public access interface mentioned earlier. Effectively, such an interface is nothing else but a well designed user interface for an interactive system. An example of the general requirements of such an interface are given in [6]. These are:

(a) free the user from being an expert at using the system in order to perform a task
(b) to be easy to learn, understand and use from a subjective user point of view
(c) to be adaptable to the user - to help solve problems
(d) to allow user control of the dialog
(e) to offer means to accomplish or allow new means as well as new tasks to be incorporated
(f) to be reliable and available.

Extending the above requirements to introduce Voice I/O while considering the application areas which have been discussed, the following Voice I/O application layer services are obtained:

(i) to free the user from being an expert ... :

To achieve this goal an interface must provide clear and descriptive instructions upon request. Commands must allow for help at nearly each state of the interface operation. Voice is ideal for this purpose since a standard set of descriptive help messages may be made available for this purpose. Although space requirements may grow with the degree of sophistication of the help facility, the nature of the help function will rarely be changed so the messages need only to be installed once. Information should also be available about the current state of the interface and what may be done at this point. It would be useful as well to maintain a history of the users progress through the session to allow back-tracking.

(ii) to be easy to learn, understand and use ... :

Since this a purely subjective measure of performance, it is difficult to define how to achieve this goal methodically. One can postulate that since speech is the natural human mode of communication, system messages may be constructed in a form which could appeal to the user from a psychological point of view. Voice can convey information about a situation through cues in intonation and accent. This may be used in conjunction with the literal message to enhance the understanding of the given state of the interface. Both learning and easy use of the system come from a comprehensive set of prompts, commands and clear system responses.

(iii) to be adaptable to the user... :

Adaptability implies that the system should be capable of allowing the user to tune the interface performance to the user's capabilities. Such modes as "Novice", "Intermediate", "Expert" allow the user to take full advantage of interface capabilities without being confused by cryptic command terminologies.

In addition, an appropriately designed interface attempts to guide the user through the system with interpretable responses.

Voice command input and voice system response can serve to enhance the naturalness of the user interface, and may be configured to support user performance matching.

(iv) to allow the user to control the dialog ... :

In a voice communication the user is most likely to converse in a natural fashion (if so allowed by the interface). Thus, it is most likely that the natural tendency of the user will be to control the dialog in a conversational way. The interface must expect and attempt to support this conversational mode of interaction. This is realized through a set of dialog control commands.

(v) to offer means to accomplish tasks ... :

The means needed to accomplish tasks are generally task-oriented commands within the interface which refer to task-performing resources. These must be identified within the system and a command structure constructed to support this requirement. Voice I/O may make this structure more flexible since the task oriented instructions may be used on a higher level. Keywords may be used to allow a greater degree of command access freedom. Arbitrary points in a tree structure, for example, may be accessed through voice command interpretation. Care should be taken to preserve the conversational nature of the voice dialog.

(vi) to be reliable and available ... :

These requirements are most likely the most difficult to satisfy at present. Reliability is still a question with speech recognition systems. If an expensive system with proven reliability is employed, it must be a shared resource in the system due to cost. This raises the question of availability. Both reliability and availability are a major consideration in Voice I/O implementations.

To summarize the Voice I/O interface functions required for the Application Layer:

(a) to provide clear and descriptive instructions available upon request;

(b) to provide help responses available at nearly every state of interface operation implying a well defined set of states with an associated set of descriptive help messages;

(c) to provide information about current status of the interface and of the task being performed;

(d) to convey (or not convey) side information to the user via the voice intonation and/or accent;

(e) to adapt or be adapable to the user's capabilities;

(f) to maintain a conversational mode of operation to exploit existing human conversational traits and characteristics, while allowing the user to control the dialog;

(g) to provide verbal commands and responses which are task-oriented to allow the user to complete tasks expediently and unambiguously;

(h) to be reliable in the sense that errors made by the system are not hidden and, if possible, nondestructive;

(i) to be available to the user in a way that the user does not realize the interface is a portion of a network and is being shared.

II.4.2 The Presentation Layer

In review, the primary functions of the Presentation Layer are management of structured data entry, exchange, display and control. From a practical point of view, the Presentation Layer supports the greatest Voice I/O data handling load. It is at this level that the pre-processing of voice data occurs. The services required of this layer are embodied in a terminal voice protocol. Such a protocol specification results from the following description of required services:

(i) the layer must accept encoded voice data from an encoder and transform this data according to an agreed upon format into a raw voice message;

(ii) the layer must communicate raw voice messages to network voice recognition resources and receive back system command messages;

(iii) the layer must employ a known data structure for both voice and data - it must support non-voice operation transparently;

(iv) the layer must accept command messages passed from an application and transform them via a known format into voice response messages if required;

(v) the layer must communicate voice response messages to the network response resources (most likely local);

(vi) the layer must be able to accept varying, but known voice message formats for transformation from a given format to a required different format;

(vii) the layer must support transformation of display commands to appropriate display instructions;

(viii) the layer must maintain the status information of speech processing resources in the network.

II.4.3 The Session Layer

The Session Layer is responsible for both session administration and session dialog services to the Presentation Layer. The Session Layer in general supports the connection, maintainance of dialog and termination of a communication between Presentation Layer entities.

Services required of the Session Layer are:

(i) to initiate and establish a connection between communicating presentations entities upon request;

(ii) to maintain the session and re-establish it if an unwanted termination occurs;

(iii) to establish a session termination upon request;

(iv) to maintain a message format which identifies the source and sink presentation entities;

(v) to identify in the message format the voice attributes of the session:
  (a) determine if this is to be a voice operated session;
  (b) determine what Voice I/O resources are available (if any);
  (c) identify what priority the session messages have;
  (d) identify the message in terms of the information it contains (command, document, page, version number, etc.).

(vi) to support dynamic redefinition of terminal characteristics such as:
  (a) terminal primitives not using voice;
  (b) terminal voice equipment reconfiguration;
  (c) terminal primitives using voice.

(vii) to maintain a knowledge of the location and identity of voice processing equipment in the network.

II.4.4 Lower Layers

From the Transport Layer down, services provided by the network are effectively a transparent data transport mechanism between session entities (in association with lower layers). Voice I/O data communications is thus supported in a manner identical to any data channel, with some exceptions. For example, a digital telephone network requires considerations made with respect to (near) real-time operation. This is relevant to our analysis since integrated services are considered in the layering approach. Responsibility for maintaining real-time operation is shared among lower and higher layers. Lower layer transmission delays due to virtual circuit congestions, for example, have been dealt with through flow analyses and various data handling algorithms. The requirements for real-time voice transmission in a network have been studied [7, 8, 9], and references to existing systems are made later to accommodate the data transport requirements of the Voice I/O functions. Real-time speech communication is more an issue for integrated services support rather than a direct requirement of Voice I/O.

III IMPLEMENTATION OF VOICE I/O IN TELIDON VIDEOTEX

At this point, the basic functional requirements of the Voice I/O interface have been discussed from a network architecture point of view. These requirements, as well as several other implementation issues which are now presented, form the basis for establishing directions for the incorporation of Voice I/O into the Telidon system concept. Outside the network layer requirements, consideration must be made to the adoption of appropriate speech processing technologies and implementations. The criteria governing the selection of a given speech technology are:

(a) speech quality,
(b) coder complexity/cost,
(c) code generation and compatibility,
(d) storage requirements (bit-rate).

(a) Speech Quality:

Speech quality measures are based on subjective testing of the speech output from specific coder implementations. A completely objective speech quality measure has yet to be defined, although methods have been specified to serve as guidelines in assessing the subjective quality of a given speech encoding scheme [14].

In relation to Telidon Voice Output, speech quality assessment is critical to the selection of a given technology. "User acceptance is fundamental to Videotex" [15]. This statement implies that great care must be taken in deciding just what is "acceptable" quality speech. It is difficult to imagine that users would find less-than-commentaryquality voice eminating from their television sets acceptable. On the other hand, people have daily experience with telephone quality speech.

It is herefore important to consider some way of defining a suitable quality level for speech applied to Telidon Voice I/O, keeping in mind the experience the user has with existing broadcast material. In studying the coding requirements for ISDN standarization, CCITT are currently considering several catagories for speech and sound digitization [[31,32]. Among these is the coding of speech with wider-than-telephone bandwidth (4.5 - 7 kHz). Considering that marked quality improvement is noted for encoded speech with a 7 kHz bandwidth [33,34]| and that both speech and music are currently broadcast within a 5 kHz bandwidth on AM radio, a bandwidth in the range 5 - 7 kHz is likely to suit Telidon Voice I/O quality requirements.

(b) Coder Complexity/Cost:

In most cases, the cost of coder implementation is directly proportional to coder complexity. Conversely, coder complexity is inversely proportional to coder data rate. Thus, a compromise between coder cost and data rate savings must be made. With the decreasing costs of digital hardware, coder complexity may increase without major increases in implementation costs. This does not necessarily imply that these relative costs have decreased sufficiently in an absolute sense to be used in Telidon terminals. Telidon terminal costs must suit mass market prices with only small initial sales. As such, terminal costs must be absolutely minimized. If expensive, add-on features such as voice may not be easily accepted.

On the other hand, if the terminal voice capability is compatible with other services available to the user via an integrated services network, more complex coders may be viable. As well upward compatibility to future schemes should be considered when more complex hardware becomes cost-effective.

(c) Code Generation and Compatibility:

The generation of voice data for storage with picture data in a Telidon system must be accommodated. Information Provider systems exist for graphics "page" generation and similar system must be designed for the integration of both voice and data. As with Telidon PLP, initial specification of voice protocols and page formation protocols will change with time. An encoding scheme or encoder type must be specified to allow upward compatibility of data with new systems.

As well, encoding scheme inter-compatibility must be maintained, requiring a standardized approach to speech data generation for storage purposes. Any information provider terminal using one of a set of accepted encoding schemes must be compatible within the Telidon system.

(d) Storage Requirements (Bit-Rate):

Storage requirements, which directly relate to coder bit-rate, are important when considering data base size. Clearly, if fewer bits are needed to represent speech information, less space for "page" storage is required. This would seem to indicate that the lower the coder bit-rate the better, to minimize storage requirements. This in fact may not be the case. Since speech quality is key factor in the acceptance of Voice I/O, care must be taken in specifying the degree to which bit-rate gains override speech quality.

The amount of speech associated with a given page of picture information should vary significantly with content and application. In the same way that film documentaries are made where visual images complement verbal information or vice versa, Telidon speech is not likely to simply reiterate what appears on the screen. As such, research must be done to determine realistic average values for the number of seconds of speech expected per frame based on well constructed picture-voice presentations.

A speech technology should not be selected in isolation from the system within which it will operate. The following section describes a terminal subsystem which includes voice capability. The terminal is represented by a block diagram and the voice module is allowed to take on several configurations.

II.1 A Voice I/O Terminal Subsystem

Figure 3 shows a block diagram of a terminal subsystem suitable for a variety of voice applications. A voice codec is shown, directly interfaced to a CPU, one of two in the system. This allows preprocessing of the digitized voice independent of the PLP operation, for example. The two processors run either out-of-sync (i.e., they are of the type that access the bus on a certain phase of the system clock such that one may run slightly out of sync from the other without bus conflict) or in a master-slave relationship. Other configurations are possible such as:

1. a CPU for voice and data handling with a custom chip set for PLP decoding,
2. custom chip-sets for voice and for PLP decoding,
3. entirely integrated PLP decoder with speech capability which conforms to a protocol standard.

The use of two CPUs for the processing of data within the terminal subsystem is representative of the computational requirement of an intelligent terminal. A highly flexible interface using voice would entail concurrent voice/picture presentation. If voice occurred sequentially with picture presentation; i.e., voice message-picture - voice message, full advantage of the interface would not be realized (since both audio and visual information can be assimilated simultaneously). Hence, processing of voice should be concurrent with picture presentation.

An initial implementation of the voice module may be as shown in Figure 5 with a CPU, memory and a codec. A more general purpose approach may be to use a real-time signal processor for implementation of the voice codec. The signal processor should have the capability of realizing several coding technologies based on a programmed configuration. This would allow dynamic reconfiguration of the codec.

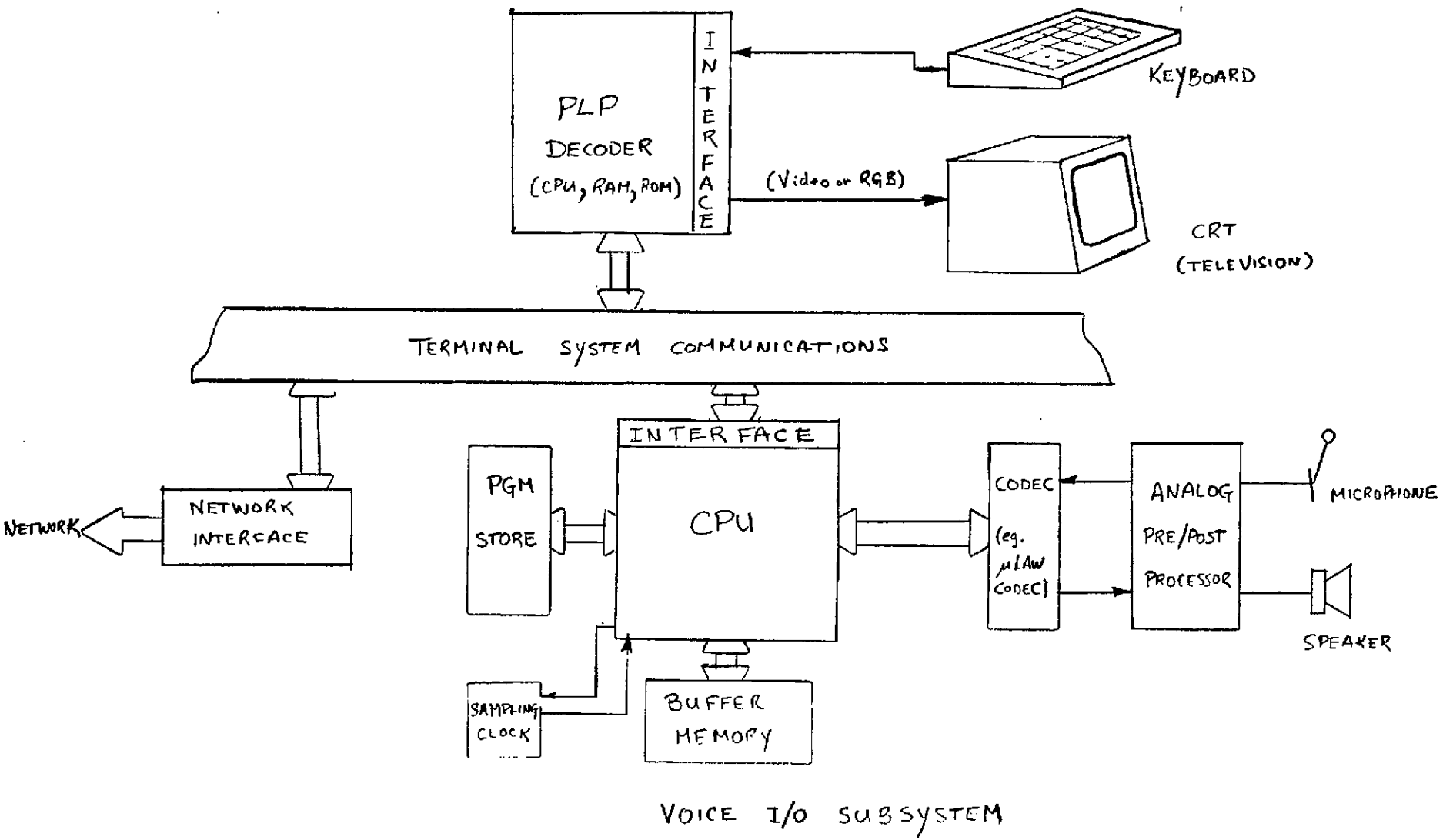
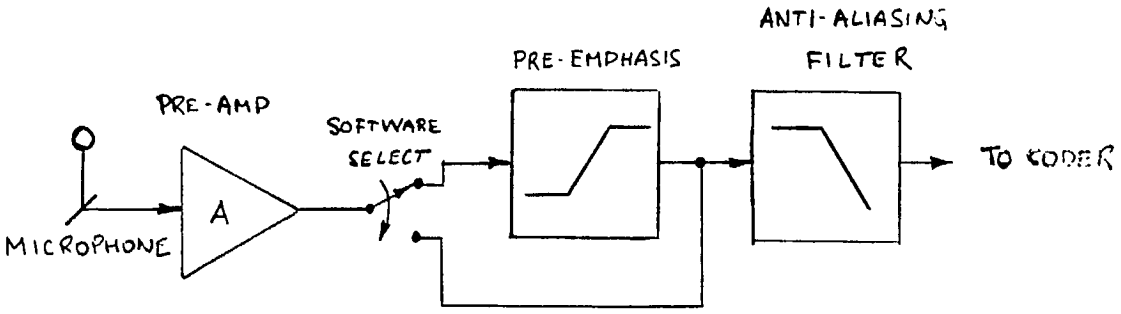
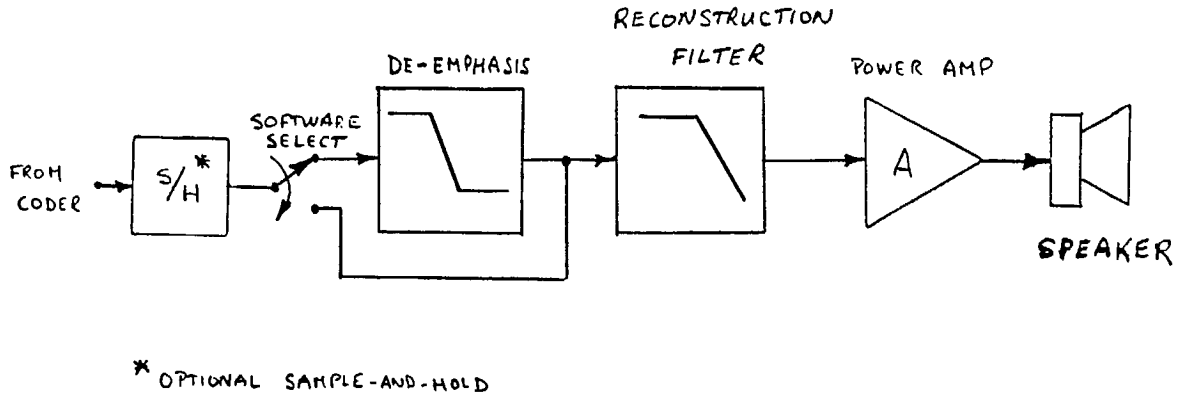


FIGURE 3 - TERMINAL BLOCK DIAGRAM



A) PRE-PROCESSOR (WITH SELECTABLE PRE-EMPHASIS)



B) POST-PROCESSOR (WITH SELECTABLE DE-EMPHASIS AND OPTIONAL SAMPLE-AND-HOLD)

FIGURE 4 - ANALOG PROCESSING MODULE

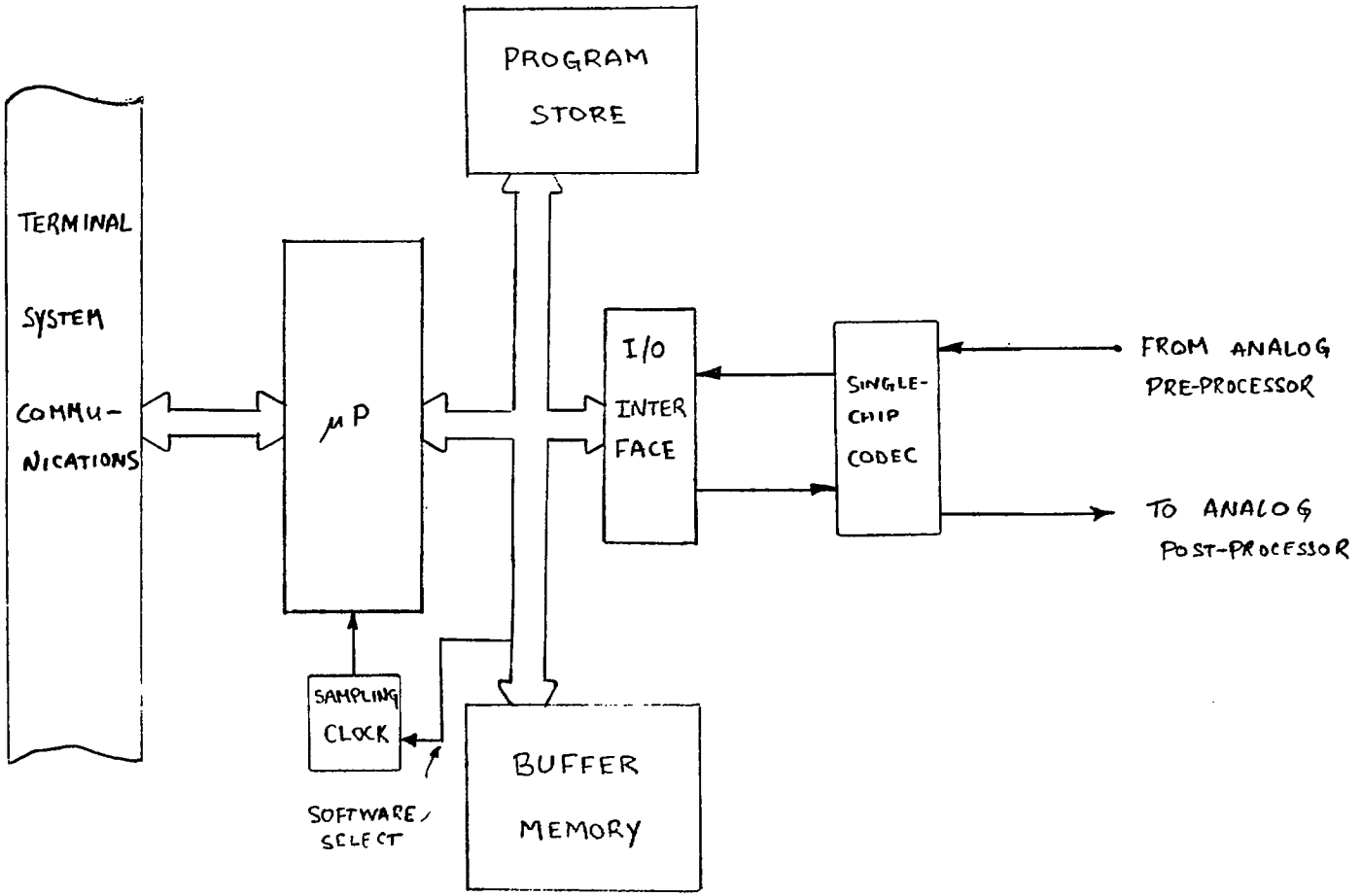


FIGURE 5 - GENERAL-PURPOSE TERMINAL VOICE MODULE BASED ON A μP AND SINGLE-CHIP CODEC

III.2 Selection of the Codec

Discussions on the relative merits of different speech coders can be found in the literature [16,17].

As discussed earlier, wider bandwidth encoding of speech is required for Telidon Voice I/O. Encoding 5 kHz bandwidth speech at 12 kHz (2 kHz greater than Nyquist rate) results in a bit-rate of 96 - 144 k bits/s for 8 - 12 bit PCM. The proposed ISDN basic access B - channel is set at 64 k bits/s [31]. As a result, straight forward PCM encoding is not applicable unless rates greater than 64 k bits/s are accommodated. This may be possible if used in wideband networks. In attempting to accommodate the 64 k bits/s capacity, studies have shown that the following techniques are applicable [33,34,35] :

- ADPCM (fixed or adaptive predictor)
- Split Band ADPCM/APCM
- TDHS - PCM (time-domain harmonic scaling - PCM)
- Sub Band Coding

As well, CCITT has heard proposals from European Administrations on the use of Adaptive Delta Modulation (ADM)(Continuously Variable Slope and Nearly Instataneously Companded) for wider band voice encoding.

Of the above mentioned coding schemes the least complex to implement is ADM. Several commercially available CVSDM codecs exist (Motorola MC3417/18, Harris HC-55516/55532, CMA FX 209) and additional development of single-chip devices is taking place [37,38] . If such a device were used, the structure in Figure 5 could be employed with the analog pre/post processor given in Figure 4. This structure does not allow for coder reconfiguration, but provides an inexpensive solution if ADM were used exclusively. This is not expected to be the case, however.

An alternative to the implementation structure in Figure 5 is shown in Figure 6 (using the analog processor of Figure 4). In this structure, the bulk of the module is made up of two digital signal processors. Examples of such devices are the Bell Labs DSP, the Intel 2920 and the NECμPD7720. Speech coder implementations have been realized using the DSP [18,19] and the μPD7720 [36]. Using such a device, if cost effective, is superior to the implementation in Figure 5. The architecture and capabilities of a digital signal processor are taylored to implementing signal processing algorithms efficiently and thus allow more complex algorithms to be realized on a single device. A drawback of some devices (DSP, Intel 2920) is that the device contains ROM program stores, which does not allow dynamic reconfiguration. If a digital signal processor is fabricated for Telidon purposes, this drawback could be rectified using other memory technologies (RAM, EEROM).

Associated with the digital signal processors may be a single chip microprocessor for code conversions and simple mapping schemes. Such a device is optional, but would be used if no other data handling capability were available within the terminal.

Specific coding and adaption schemes have not been suggested in this section since further research must be done to determine which algorithms will suit the application of Telidon Voice I/O. Interactive conversational mode systems with both voice input and output have been implemented [20,21] and general approachs for serving large consumer groups have been studied [22, 23,24]. In the former case, a waveform coding technique (ADPCM) was used for voice response. In addition transformations exist between different encoding schemes such as PCM-to-ADPCM directly in the digital domain [25] This allows for a wide variety of coder

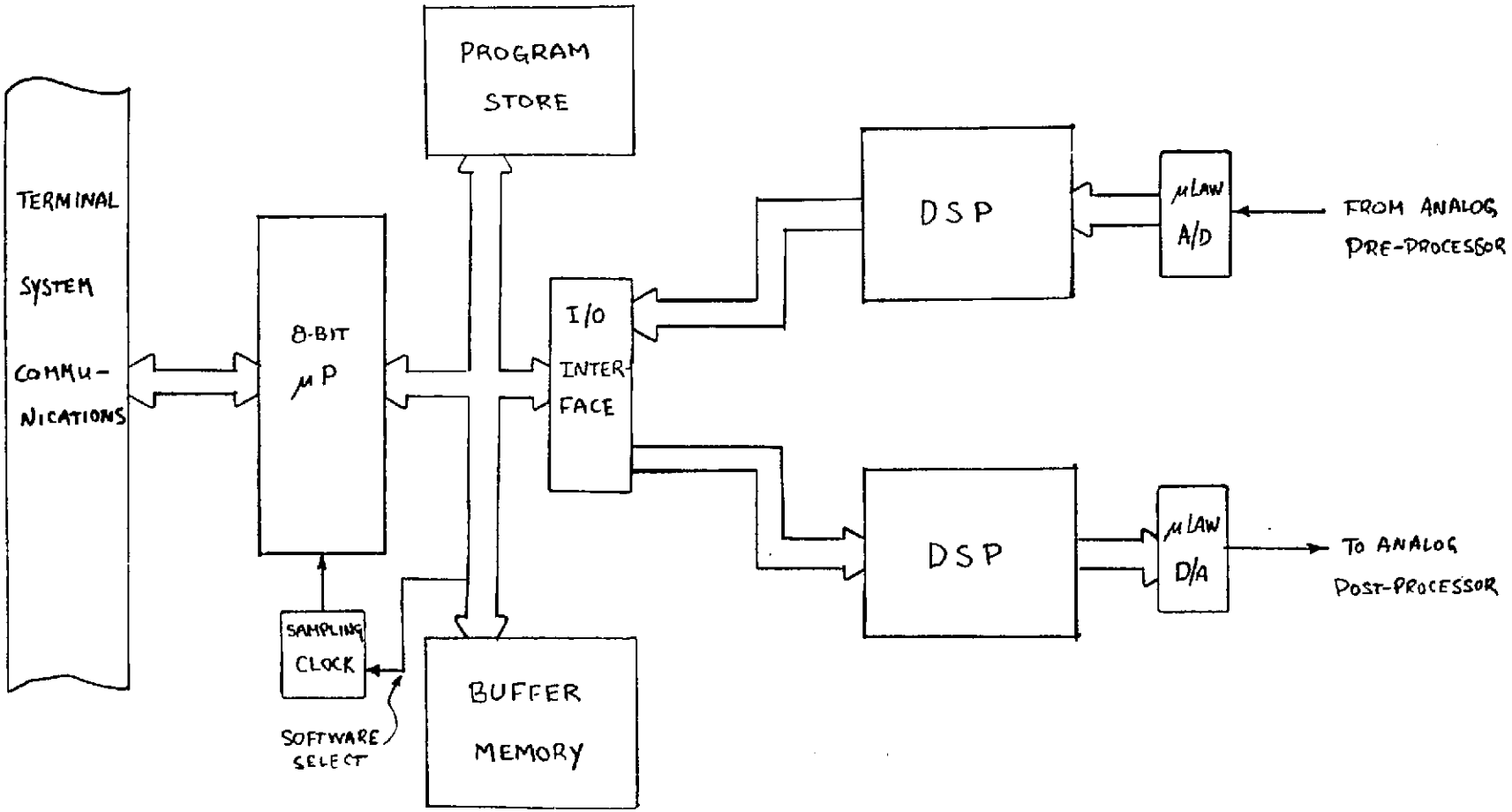


FIGURE 6 - TERMINAL VOICE MODULE BASED ON DIGITAL SIGNAL PROCESSORS

implementations and variations. A key aspect of this is the need for establishing a set of standard approaches to encoding voice and embedding this within the Telidon standards. From a protocol point of view the technique should remain transparent to the protocol elements i.e. any scheme will operate in the Telidon environment.

Having shown a possible terminal sub-system implementation with a waveform coding scheme, consideration must be made regarding voice recognition requirements.

III.3 Voice Recognition

The voice recognition aspect of the Telidon user interface is significantly more complex than voice response. Voice recognition techniques are continually being refined and developed, yet progress toward an inexpensive universally applicable system is very slow.

A comparative review of voice recognition is given in [26]. An implementation which resembles the type of system that might be used within a Telidon environment is given in [20].

A prime consideration of a voice recognizer design is the task-specific nature of the recognition scheme. Typically, the more task-specific the recognizer vocabulary, the more feasible the implementation of the system [26,20,23]. By using information from context and semantics the predictability of responses is enhanced significantly. As a result less computationally elaborate systems may be implemented.

Even though the required computations may be decreased through careful design of the user-machine dialog, some practical implementation problems exist.

The digitization of speech for recognition purposes is performed at the terminal end with the same coder used for other voice communication. Care must be taken to ensure that the coder resolution is sufficient to allow processing of the signal for recognition purposes.

Secondly, noise contamination of the signal (partially quantization noise) from external sources such as air conditioning, background conversation, room reverberation, etc. is a major consideration in the implementation of the recognizer. Since the use of the recognizer will be in various environments where there may be typical background noise present, the correct selection of microphone, terminal placement and recognizer noise normalization processing are essential for consistant performance of the recognition system. Research into characterizing the physical environment for Telidon use, selection of appropriate headset microphones (possibly noise-cancelling) and implementation of various noise normalization processes in the recognizer must be performed to fully specify the requirements for Telidon voice input.

The reliability of the voice recognizer (i.e., recognition accuracy) must be determined from several points of view. The following factors influence the feasibility and performance of a voice recognizer [26]:

i) type of speech input: connected or isolated,
ii) acceptability of response time,
iii) acceptability of recognition accuracy,
iv) size of the vocabulary,
v) speech style of the input: conversational or artificial
vi) nature of the user interface: graceful error recovery, natural interaction with the user.

It has already been noted that a conversational style is desired for the speech input and that both graceful interaction and error recovery are desired. Vocabulary size may vary according to the task and may be structured such that a given state in the interface has a certain number of expected or allowed responses, thus decreasing the number of possible responses to be recognized at any given time. Recognition accuracy and the acceptability of errors is related to the error recovery strategies, but behavioural studies are needed to determine the tolerance of a user to errors in recognition and to determine appropriate recovery strategies.

It is expected that if a large group of untrained users is to operate the recognizer (as is true with Videotex), connected speech will be the typical mode of input. This will involve more complex processing algorithms and may result in significant delays before commercial systems become available within the Telidon system.

Response time is affected by two major factors: i) the time required to perform the recognition; and ii) the time required to pass the voice data to the recognizer and receive a response (this may even be further compounded if voice response is included in the interaction). Due to the high cost of speech recognition equipment which satisfies to a degree the requirements of the Telidon Voice I/O interface (from $20,000 to $100,000), the incorporation of voice input must be initiated through shared use of an expensive system, with cost sharing by the users. This imposes a further constraint on the recognizer that it must be capable of time-sharing. As well, this adds to response delays if a large number of users are accessing the system. An initial amortization of the cost may only come with the minimization of other system utilization costs: terminal costs, connection

costs, on-line costs or data transmission charges. This is a significant drawback in terms of when such systems will be made available to the general public. Since Videotex and ISDNs are still emerging technologies, implementation of voice input is not expected to be commercially available until these technologies are firmly established.

IV. CONSIDERATIONS FOR INTEGRATED SERVICES NETWORKS

Incorporating Voice I/O into Telidon Videotex has been discussed with integrated services in mind, but direct reference to this potential for a broader scope of services will be made in this section.

A single distribution medium for Videotex is not likely to exist for some time. With controversy regarding the use of existing telephone networks vs Cable TV networks [2] and with the emerging ISDN and fibre optic technologies, the medium for distributing an integrated set of services will be developing over the next several decades.

The first indications of a higher speed channel for interactive service distribution are the ISDN and the Cable TV networks. The ISDN would essentially provide a 64 k bit/s data channel and a data and signalling channel of 8 kHz on the public telephone network. The initial phase of such a system is scheduled to be in service in the UK in June, 1983 [12].

The added bandwidth of such a network (over that of the existing telephone network) makes a Voice I/O system more appealing from an added service point of view. Services may be included in such a network through the use of Value-Added services implemented via operational modules [27] . These modules may be incorporated into the network at an ISDN switching

centre through a standard interface. In this way, several new services may be implemented throughout the network in a central or distributed fashion.

The ISDN seems to be the first step to expanding the capabilities of the telephone line to supply data communication and information retrieval services to the home.

Other than the ISDN, Cable TV networks are a second possible carrier of integrated services. Until recently the majority of cable installations have been one-way tree-like systems. This type of network structure is not suitable for interactive services and must be modified to provide a return path to the head end. Such a modification is costly and it is likely that such a system will take longer to implement. On the other hand, in an attempt to compete with telephone utilities, cable companies are considering such a scheme as indicated by the adoption of two-way systems for new installations. With the broadband capabilities of cable, a greater degree of service integration is possible. This would include TV reception, Videotex, Digital Audio and Digital Telephony. The degree of complexity of subscriber equipment would increase dramatically as compared to the simple TV set and telephone now used in the home. Required terminal capabilities and development trends are described in [28]. Incorporation of Voice I/O into a cable system would follow similar analysis to that presented in this study.

To further increase the information capacity of a public distribution network, fibre optic technology may be incorporated into the cable system. There is some question as to how feasible the implementation of such networks will be within the 1990 time frame. Arguments have been made for and against the feasibility of optical fibre plants [29,30].

The implications for Voice I/O incorporation are mostly from the terminal point of view. In general, the terminal will have to be highly integrated to keep costs down. In addition, different encoder approaches may result from an extended bandwidth. The increase in bandwidth of the system still does not warrant the use of high bit-rates, since storage capacity is a major factor related to bit-rate.

V. CONCLUSIONS AND RECOMMENDATIONS

The results of this study may be summarized by the following conclusions:

i) The use of the OSI model for structuring an approach to Voice I/O incorporation into Telidon clarifies the protocol requirements for the three upper layers. As well, the analysis of user requirements for foreseeable Telidon application groups yields a more general specification of the Telidon Voice I/O interface requirements from a hardware and software point of view.

ii) Selection of appropriate technologies for voice processing is directly dependent on the application requirements.

iii) The specification of voice processing technologies is required in the following areas:

1. voice digitization, storage and reconstruction hardware and algorithms,
2. voice recognition algorithm selection,
3. voice storage associated with Telidon picture information:
   a) data base structure,
   b) concurrent data handling within the terminal,
   c) Information Provider data entry system,
4. integration of Voice I/O into an Integrated Services Network.

iv) The recommended implementations of voice coding are a waveform coding approach (to maintain high quality and inter-service compatibility), such as ADPCM, ADM, THDS-PCM, or Sub Band coding at 64 k bits/s rates.

v) The use of a digital signal processor for coder implementation is recommended for efficient implementation and inherent flexibility. (Cost is a major limiting factor here.)

vi) The initial incorporation of voice input should be on a user-shared basis, rather than on a per terminal basis. In this way a more expensive system which is required to support the service requirements may be used, and the costs shared by the users.

Continued research is required in the following areas:

i) establishing voice protocol standards for the higher layer protocols,

ii) integration of a terminal voice module which will recognize voice protocols (similar to an integrated PLP decoder),

iii) defining a suitable user voice interface which is taylored to Voice I/O information access for Telidon Videotex. This includes behavioural research into appropriate allocation of information to picture and voice,

iv) further research into voice coding and recognition techniques for large groups of users.

v) development of integrated circuits for coding algorithms suited to wider band speech coding (0 - 7 kHz) at 64 k bit/s rates.

APPENDIX A

A.1 Layer Protocol Specifications

In this study a major objective is to resolve the protocol issues associated with incorporating Voice I/O into the Telidon system. The OSI model was selected to structure an approach to specifying these issues. The specification of the protocols for the upper three layers only will be addressed in some detail. Lower layers effectively support error-free data transmission and tend not to be an issue in incorporating Voice I/O. The major thrust is toward specifying higher layer protocols especially within the Presentation Layer, where many details on terminal protocols need to be resolved. At the Application Layer, effective man-machine interface algorithms must be specified with Voice I/O as a communication service in order that voice capacilities are used to the utmost.

The following protocol specifications have not been formally verified and are included to form a basis for suggested directions for Voice I/O incorporation into Telidon.

The protocol specifications follow a specification method given in [10]. Following this specification method, the input/output behavior of the protocols may be defined through a Service Specification of the protocol. Exact formats of commands and the mechanisms to convey them are defined by an Interface Specification.

Prior to embarking on a service specification of the higher layer protocols, a brief description of the protocol structure, the manner in which a communication service may be defined and the internal structure of a protocol layer are described.

As mentioned above, the input/output behavior of a protocol is defined by a service specification. Typically, a service specification may be based on a set of service primitives. The execution of a service primitive is associated with an exchange of parameter values between two entities across a layer-layer interface. This is to say that two entities in a layer communicate via the layer below them through the execution of service primitives which in total comprise a service specification. In general, service primitives have constraints as to the order of execution and the values of the parameters exchanged during execution. These constraints may be global (referring to other users) or local (referring to the immediate user of the service).

A protocol characterizes the interaction of entities within a layer (via the lower layer services) to provide the services of that layer. Thus a protocol specification for a layer must define the operation of each layer entity in response to:commands from the users of that layer's services, messages from entities within that layer,and actions initiated internally.

A variety of methods for formal protocol specification are given by Bochmann and Sunshine [10]. Since a goal of this study is to place the protocol issues of Voice I/O incorporation into Telidon into perspective, the protocol specifications are intentionally narrative.

A.2 Presentation Coding Structure

A.2.1. Test Coding: (7 bits)

The ASCII code table is given in Figure 3,3 of [11]. ASCII (American Standard Code for Information Interchange) has been accepted as an internationally accepted character coded text assignment code (ANSI, CSA and ISO). As can be seen in this figure, the first two columns correspond to control character assignments while the remaining columns represent character codes. The control character set, called CO, may be subdivided into transmission control characters (the shaded areas), Format Effector (FE) characters, Information Separator (IS) codes and Code Extension (CE) codes. In addition, the Bell (BEL) code is present. As a rule, transmission codes are not used by presentation layer processes.

Other than the transmission codes, the FE, IS and CE codes are available in the PLP. NOTE: CAN may be recognized by PLP as a reset if sent by lower layers.

The ASCII code table is mentioned in this discussion since it becomes the default graphics set in the graphics repertoire of the code extension approach defined by ISO Standard 2022, and represents a standard code table format (for 7-bits). As in the ASCII code table, a general format for an in-use code table may be defined, where the first two columns contain a

control set (C set) and the remaining columns contain a set of graphic character assignments, called a G set. As mentioned, the default G set, G0 is the ASCII character set. The C0 set is defined as a harmonized CCITT control set.

Currently, two active C sets are defined (C0, C1) and four active G sets are available (G0, G1 G2, G3). Figure 3.10 of [11] shows code extension using 7-bit character codes where the 8th bit is used for parity.

In an 8-bit environment, the most significant bit may be used to select between two G sets, GL and GR, within a single in-use table. As well, both C0 and C1 are present in the table simultaneously. Figure 3.11 & 3.12 of [11] denote the code extensions available using 8-bit code assignments.

In specifying Voice I/O protocols within the PLP coding structure the approach taken is to define a G set which contains protocol primitives which implement the functions of the Presentation Layer defined earlier. As the specification proceeds, existing PLP primitives are described, where applicable, as analogous operations to those specified for Voice I/O. They are analogous in the sense that existing PLP primitives, namely Picture Description Instructions (PDIs) perform functions which may be similar in a control sense to Voice I/O protocol elements, but refer to screen data representations not voice data representations.

A.3 A PLP Voice Protocol

A PLP Voice Protocol approach, loosely speaking a set of Speech Description Instructions (SDIs), is now defined according to Presentation Layer Voice I/O requirements. Initially only macro-instructions are defined for major functions.

The specification begins with the Presentation and Session Layers which will be initially treated to some extent as a single virtual terminal protocol. At this point the existing Telidon Videotex Presentation Level protocol is introduced.

A.3.1 Telidon Videotex Presentation Level Protocol

This section deals with the specification of the Presentation Level functions for Voice I/O given earlier in this study within the existing Telidon PLP (Presentation Level Protocol). Rather than describing Telidon PLP in detail, reference is made to CRC Technical Note #709-E [11] where Telidon PLP is completely specified. Only a brief introduction to the PLP structure is given to allow the Voice I/O specification to be made in context. This specification is one of many possible functional implementations of Voice I/O Presentation Level functions and is to serve as a proposal yet to be verified.

A.3.2 Constructing a Voice Message to be Recognized

In this case, it is assumed that a voice codec is available to encode speech at some defined sampling rate. Assume that hardware is also available to detect the use of the microphone (either a push button or a voice activated switch). As voice data is generated from the speech, a pre-defined passage length maximum is reached. It is preferable to delimit passages with natural speech pauses, but this is not always possible. The variability of speakers may prevent this type of segmentation.

Upon completion of the acquisition stage of the process, the raw speech data must be identified as such.

This may be implemented using a macro-intruction option, similar to the MACRO-PDI PLP opcode. If it has been established that the terminal supports Voice I/O, the application process may initialize the terminal with a transmit macro that defines a portion of buffer memory in the terminal. This buffer would be accessed by the terminal during raw speech acquisition after which the transmit macro would be executed. Data within the transmit macro would identify the message as raw speech data. The execution of this macro would result in a raw voice message to be transmitted to the network. The Session Protocol must ensure that the speech data is directed to the speech recognition resources but the Presentation Protocol must specify the destination. This should be conveyed via the macro since the Session layer will maintain node addresses for the distributed voice processing resources.

For convenience let this macro be named "SENDSPEECH". SENDSPEECH satisfies the required functions:

(i) format raw speech data into message format,
(ii) communicate raw data to speech recognition resources,
(iii) maintain some knowledge of the speech processing resources of the system.

A.3.3 Receiving a Voice Response Command

Converting a voice response command to the network voice response primitive must be performed in the case where voice data is not sent from the application process yet the Voice I/O operation is intentionally in use. As an example, a Dynamically Redefinable Character Set (DRCS) is defined where a set of messages are stored as a G set repertoire. In this case, a macro may be defined to access the alternate G set, extract the message from the set and store the data in an output buffer. At this point a transmit macro may be used to send the message to the voice response resources (which may be local).

In this case opcodes are needed for the stored messages to be transmitted to the voice response resources and for the access to the DRCS upon receipt of a command.

For convenience these will be called "MSGDUMP", "DEFMSG", "CALLMSG" where

(a) MSGDUMP defines the transmit macro for transporting the voice response data to voice response resources,
(b) DEFMSG defines the command for downloading the voice response data to be stored as a DRCS, and
(c) CALLMSG defines the command for translating the command following this opcode to a G set reference for accessing the DRCS voice response data.

These three macros satisfy the required functions of:

(i) accepting application command messages and transforming them into voice response messages, if required.
(ii) communicating voice response messages to network voice response resources.

A.3.4 Translating from Format to Format

In the case where the data received is of a format which is not immediately compatible with the existing speech response hardware configuration,

the transformation from one known format to another is necessary. By supporting such a function, the protocol becomes less susceptible to variations in information provider encoding techniques.

In order to transform one format to another, there must be either an equation relating the two, or some unique mapping of one onto the other. If this does not exist then it may be necessary to reconfigure the codec.

In the case where a transformation exists (e.g., Adaptive Delta Modulation-to-PCM) a macro command "MAP" is used. In the case where no such transformation exists, a different command should be used, similar to the initialization procedure, where the terminal is configured.

MAP satisfies the required functions:
transform varying voice message formats from one to another on request (if possible).

A.3.5 Terminal Reconfiguration

Although the state of a terminal, from a hardware point of view, should not be a concern at the Presentation Level, consideration must be given to the variation of encoding technologies which may be employed for voice communication within the network. By incorporating a capability for terminal reconfiguration without specifying the reconfiguration explicitly (i.e., "change this subroutine call to that one, etc.") a much needed requirement for flexibility with respect to emerging technologies is satisfied.

This may be achieved through a macro command "CONFIG", which is followed by a keyword indicating the desired coding technology. The knowledge of which technologies are supported by the terminal (out of a finite known set of standard schemes) is maintained at the Session Layer.

CONFIG may be made to resemble the DEF DRCS PDI where the DRCS is in fact the new configuration parameters required for the terminal. The CONFIG command would tell the terminal that such information is following and the new "G set" would be loaded into the codec thereby reconfiguring it. Alternately, the data following the CONFIG command may represent a program of an algorithm for reconfiguring the terminal speech codec.

Thus, CONFIG supports the function:

instruct the terminal to support different coding schemes for voice through reconfiguring the codec (if possible). If not supported, the above information is ignored.

In general, if no speech services are supported by the terminal, operation as a PLP terminal for Telidon must be supported. The Session Layer should determine terminal status and capabilities as the session is established and support either PLP and/or Voice I/O.

Summarizing the macro instructions defined: SENDSPEECH, MSGDUMP, DEFMSG, CALLMSG, MAP and CONFIG. These macros must now be specified in terms of terminal primitives which become the protocol primitives of the presentation level Voice I/O protocol.

A.4 A Session Layer Protocol

A.4.1 Session Establishment and Negotiation

In the establishment of a session other than determining whether two presentation entities are available for a connection, certain information must be passed from session entity to session entity for session administration purposes. This is considered to be a negotiation process. The following negotiation services must be implemented via protocol elements. Voice Input/Output session information is used to:

(a) establish a Voice I/O session,

(b) identify and match Voice I/O terminal capabilities,

(c) identify and establish connection with Voice I/O network resources,

(d) establish message format,

(e) establish and administer terminal parameter initialization.

A.4.2 Session Management

After the initial negotiation process, the remainder of the session interaction is devoted to maintaining the session until termination is requested or occurs independently. (Should the latter case occur, possible reestablishment of the session may be required of the Session Layer.)

Session maintainance requires the following elements:

(i) identify the presentation level message type: command, page, raw voice message, voice response command, etc.,

(ii) administer redefinition of terminal primitives through a renegotiation of terminal parameter initialization,

(iii) appropriately identify the destination of given message types: voice response-to-voice response equipment, raw voice message-to-speech recognizer, etc.

The Session Layer may contain the following protocol characters:

Session Negotiation:

	Sent Character	Response	Function
	CSI	RSI	- session establishment
	CSV	RSV	- voice session establishment
(in response to RSV)	CCD	RCD [0-N] [( )]	- codec definition - [·] number in brackets represents one of N coder configurations to be used in the session. Zero is reserved for no coder available (i.e., voice response only) - [(·)] optional character if [·] was zero to indicate what type of voice response device is available
		or RRC	- response sent requesting coder reconfiguration
(in response to RRC)	CCR [data]	RCR	- coder reconfiguration command followed by reconfiguration data - RCR is response that coder reconfigured
(to speech recognizer)	CCVI [addr][0-N]	RCVI	- establish connection to network Voice input resource - [·] indicates type of speech data to be received
(to speech synthesizer)	CCVO [addr][0-N]	RCVO	- establish connection to network Voice output resource
(if speech synthesizer local)	CELVO	RELVO	- engage local voice output resource
(if speech recognizer local)	CELVI	RELVI	- engage local voice

Session Management:

	CRC [0-N]	RCRC	- request to reconfigure coder to one of N schemes
(in response to CRC)	CCR [data]	RCR	- same as in negotiation dialog

Session Management (continued)

Sent Character	Response	Function
CIC [code][data]	RIC [code]	- definition of command data followed by command data - return response character with identified command code for confirmation

A.5 The Application Layer

Formal specification of Application Layer protocols is not pursued. Proper protocol specification of the Application Layer must include behavioral research which is beyond the scope of this study. In suggesting and specifying realizations for Voice I/O in the study, reference is made to research done in man-machine interfacing including voice interface applications. These references indicate possible directions to protocol specification at the Application Layer.

REFERENCES

1. Wecker, Stuart, "Computer Network Architectures", Computer, September 1979, pp. 58-72.

2. Godfrey, David and Chang, Ernest, The Telidon Book, Press Porcepic Ltd., Victoria, B.C., 1981.

3. Gecsei, Jan, Architecture of Videotex Systems, Prentice Hall, Inc., Englewood Cliffs, N.J., 1983.

4. Zimmerman, Hubert, "A Standard Layer Model", Computer Network Architecture and Protocols, Paul E. Green, Jr., Editor, Plenum Press, New York, 1982.

5. Sablatash, M. and Fitsgerald, R., "Towards the Design of an ISO OSI Layered Architecture for a Canadian Broadcast Telidon System",

6. Taeuber, D. L., et al., "A Functional Model for Interactive Systems", IEEE Proceedings of the 1982 Zurich Seminar on Digital Communications, Man-Machine Interaction, March 9-11, 1982, Zurich, pp.

7. Aoyama, Tomonori, et al., "Packetized Service Integration Network for Dedicated Voice Data Subscribers", IEEE Transactions on Communications, Vol. COM-29, No. 4, November 1981, pp. 1595-1601.

8. Cohen, Danny, "A Voice Message System", Computer Message System, R. P. Uhlig (Editor), North Holland Publishing Co., 1981, pp. 17-28.

9. Bellamy, John, Digital Telephony, John Wiley and Sons, New York, 1982.

10. Bochmann, Gregor V., and Sunshine, Carl A., "A Survey of Formal Methods", Computer Network Architectures and Protocols, Paul E. Green, Jr., Editor, Plenum Press, New York, 1982.

11. O'Brien, C. D., et al., "Telidon-Videotext Presentation Level Protocol: Augmented Picture Description Instructions", CRC Technical Note, No. 709-E, Ottawa, February 1982 (pre-print edition).

12. Clarke, K. E., "Second Generation Videotex in the United Kingdom", IEEE Proceeding of the 1982 Zurich Seminar on Digital Communications, Man-Machine Interaction, March 9-11, 1982, Zurich, pp. 107-110.

13. Lacher, S., et al., "ISDN Subscriber Loop Protocol", IEEE Proceedings of the 1982 Zurich Seminar on Digital Communications, Man-Machine Interaction, March 9-11, 1982, Zurich, pp. 77-80.

14. IEEE Recommended Practice for Speech Quality Measurements, September 1969.

15. Stewart, T., "Human Factors in Videotex", Butler Cox and Partners Ltd., U.K.

16. Flanagan, James L., et al., "Speech Coding", IEEE Transactions on Communications, Vol. COM-27, No. 4, April, 1979, pp. 710-737.

17. Holmes, J. N., "A Survey of Methods for Digitally Encoding Speech Sigmals", The Radio and Electronic Engineer, Vol. 52, No. 6, pp. 267-276, June 1982.

18. Crochiere, Ronald E., et al., "Real Time Speech Coding", IEEE Transactions on Communications, Vol. COM-30, No. 4, April 1982, pp. 621-634.

19. Goodman, David J. and Nash, Randy D., "Subjective Quality of the Same Speech Transmission Conditions in Seven Different Countries", ibid, pp. 642-654.

20. Levinson, S. E. and Shipley, K. L., "A Conversational-Mode Airline Information and Reservation System using Speech Input and Output", B.S.T.J., Vol. 59, No. 1, January 1980, pp. 119-137.

21. Rabiner, L. R., Wilpon, J. G. and Rosenberg, A. E., "A Voice-Controlled, Repertory-Dialer System", B.S.T.J., Vol. 59, No. 7, September, 1980, pp. 1153-1163.

22. Endres, Werner K., "Man-Machine Speech Communication - A Basis for New and Improved Services of the Deutsche Bundespost", IEEE Proceedings of the 1982 Zurich Seminar on Digital Communications, Man Machine Interaction, March 9-11, 1982, Zurich, pp. 348-352.

23. van Nes, F. L., "Perceptive, Cognitive and Communicative Aspects of Data Processing Equipment", ibid, pp. 259-262.

24. Quiniou, Rene and Saint-Dizier Patrick, "Man-Machine Interface for Large Public Applications", ibid, pp. 147-152.

25. Rabiner, L. R. and Schafer, R. W., Digital Processing of Speech Signals, Prentice Hall, In., Englewood Cliffs, New Jersey 07632, 1978.

26. Reddy, D. Raj, "Speech Recognition by Machine: A Review", Proceedings of the IEEE, Vol. 64, No. 4, April 1976, pp. 501-531.

27. Bocker, Peter and Gerke, Peter R., "The Integrated Services Digital Network (ISDN) and Its Use for Text and Data Communication", Computer Message System, R.P. Ulig (editor), North Holland Publishing Co., 1981, pp. 53-65.

28. Storey, John R., "Terminals for Future Home Services Received via Wide Band Networks", Technical Records of CCTA 21st Annual Convention, May 29-June 1, 1978, Montreal.

29. "Memorandum on the Technological and Cost Aspects of Integrated Distribution Plant", Bell Canada, March, 1978.

30. Jull, G. W. and Bryden, B., "New Broadband Home Services: Coaxial Cable or Optical Fibre Local Plant?", Department of Communications, Communications Research Centre, Shirley Bay, Ottawa, 1978.

31. Decina, Maurizio, "CCITT Activity on Signal Processing for Integrated Services Digital Networks", IEEE International Conference on Acoustics, Speech and Signal Processing, Paris, 1982, Vol. 1, pp. 5-10.

32. Maitre, X. and Aoyama, T., "Speech Coding Activities Within CCITT: Status and Trends", ibid, Vol. 2, pp. 954-959.

33. Bertorello, Luciano, et. al, "Broadcasting-Quality Transmission of Audio Signals at 64 kbps", ibid, Vol. 3, pp. 1972-1975.

34. Combescure, P., et al, "ADPCM Algorithms Applied to Wideband Speech Encoding (64 k bits/s, 0-7 kHz)", ibid, Vol. 3, pp. 1976-1979.

35. Johnston, I.D.; Goodman, D.J., "Digital Transmission of Commentary Grade (7 kHz) Audio at 56 or 64 k bits/s", IEEE Transactions in Communications, Vol. 28, Jan. 1980, pp. 136-138.

36. Feldman, J.A., et al, "A Compact, Flexible LPC Vocoder Based on a Commercial Signal Processing Microcomputer", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 31, No. 1, Feb. 1983, pp. 252-257.

37. Iric, K. et.al, "A Single-Chip ADM LSI Codec", ibid, pp. 281-287.

38. Tanaka, F., et al, "C^2MOS Speech Synthesis Systems", ibid, pp. 329-334.

cover design: Diane Weselake, Instructional Media Centre