ETSI TECHNICAL REPORT

ETR 131

June 1994

Source: ETSI TC-TE

Reference: DTR/TE-04012

ICS: 33.020, 33.040.40

Key words: TE, stored voice services

Terminal Equipment (TE); An investigation into the need for standardization in the area of stored voice services

ETSI

European Telecommunications Standards Institute

ETSI Secretariat

Postal address: F-06921 Sophia Antipolis CEDEX - FRANCE
Office address: 650 Route des Lucioles - Sophia Antipolis - Valbonne - FRANCE
X.400: c=fr, a=atlas, p=etsi, s=secretariat - **Internet:** secretariat@etsi.fr

Tel.: +33 92 94 42 00 - Fax: +33 93 65 47 16

Contents

Foreword

This ETSI Technical Report (ETR) has been produced by the Terminal Equipment (TE) Technical Committee of the European Telecommunications Standards Institute (ETSI).

ETRs are informative documents resulting from ETSI studies which are not appropriate for European Telecommunication Standard (ETS) or Interim European Telecommunication Standard (I-ETS) status. An ETR may be used to publish material which is either of an informative nature, relating to the use or the application of ETSs or I-ETSs, or which is immature and not yet suitable for formal adoption as an ETS or an I-ETS.

Summary

This ETR describes the results of work in response to a CEC mandate for a study on the requirements for standards in the Audiotex domain. When initial studies identified the fact that "Audiotex(t)" tends to have different meanings in different countries the CEC accepted a proposal to use the term "s**tored voice services**" to describe the field of work.

It was found that there was much confusion in the naming and description of the various stored voice services on offer but in spite of this, stored voice services are an active and growing market. The growth of the stored voice services market is not being inhibited by the lack of standards as there are a number of voluntary standards writing fora, comprising manufacturers and users, which are working to ensure that standards are available in advance of the development of the market.

It is clear that the usability of any given stored voice service has a significant effect on its potential for failure or success in the marketplace. The convenience of use can often be enhanced by interaction with other messaging services.

Future new techniques can be expected to have an effect on the manner in which voice services are presented to the customer and on the development of the market. These developments may be restricted in some areas by the non-uniform distribution of supporting communications facilities (such as the availability of Dual Tone Multi-Frequency (DTMF) signalling) throughout the European marketplace.

A number of recommendations resulted from the study.

Recommendation 1

To avoid confusion in the description and specification of stored voice services it is recommended that:

The use of the term stored voice services should be encouraged and the term Audiotex be deprecated in the field of standardization.

Recommendation 2

In order to encourage companies in Europe to take part in voluntary standards writing fora, it is recommended that:

Support should be given to participation in voluntary standards fora that are relevant to stored voice services.

Recommendation 3

In order to assist purchasing decisions in an open and competitive market, it is recommended that:

Methods should be developed to quantify the usability of stored voice services.

Recommendation 4

To assist the growth of a free market in stored voice services it is essential to improve the consistency in their operating procedures. Therefore:

Guidelines should be developed for the user procedures and dialogue in stored voice services, especially that for entering alphabetic characters on telephone sets with letters of the alphabet assigned to numbers.

Recommendation 5

To improve the uptake in the use of stored voice services, it is recommended that:

Guidelines should be developed for the preparation of introductory and user procedural text and voice messages for use in stored voice services.

Recommendation 6

In order to assist manufacturers to make more effective use of automatic voice recognition:

Support should be given for the preparation of guidance on automatic speech recognition for stored voice services.

Recommendation 7

So as to encourage the provision of multi-lingual stored voice services:

A standard for language identification codes should be developed.

Recommendation 8

To raise awareness of, and to give better information on, the design of user interfaces, it is recommended that:

A suitable guidance document should be prepared for the design of the user interface for stored voice services.

1 Scope

1.1 General

This ETR discusses the whole range of stored voice services, identifies their various characteristics and components and shows the interactions between stored voice services and other services such as facsimile, Videotex and multimedia.

It identifies existing standards and recommendations relevant to these services, analyses the need for new standards and makes recommendations for future action.

1.2 Background

The objectives of the study covered by this ETR originally were to define an Audiotex service, identify existing standards and to analyse the need for new standards.

The original Commission of the European Communities (CEC) mandate [1] described the work programme as a study on the requirements for standards in the Audiotex domain. The purpose of the study being to determine "whether there is a need to establish a work programme for what were described as "telematic services" which use standard telephone equipment to convey information to users. They can be interactive or non-interactive".

The terms of reference laid down to fulfil this mandate were to define an Audiotex service and its relation to other services (e.g. teleaction), identify existing standards/recommendations relevant for this service and to analyse the need to develop standards for:

- a description of an Audiotex teleservice;

- terminals to support an Audiotex (service) and or Audiotex applications;

- the use of Dual Tone Multi-Frequency (DTMF) signalling for accessing an Audiotex service and/or Audiotex applications;

- the use of other signalling methods than DTMF to access an Audiotex service and/or Audiotex applications.

Initial studies of Audiotex identified the fact that "Audiotex(t)" tends to have different meanings in different countries and suggested the use of the term "**stored voice services**" as a generic term. It was also found that there was no practical or engineering reason to treat the many possible voice services in different ways when considering the need for standards.

It was therefore proposed, and accepted by the CEC, that the ETR should discuss the whole range of such stored voice services, identifying their various characteristics and components, and showing the interactions between stored voice services and other services such as facsimile, Videotex and multimedia.

2 References

For the purposes of this ETR, the following references apply:

[1] Commission of the European Communities. Committee on Regulations and Standards (1991): "Study on the requirements for standards in the Audiotex domain".

[2] Electronic Publishing Services (Publications) Ltd., (1991): "The Audiotex information services market in Europe".

[3] National Economic Research Associates (1991), CEC: "Study of the application of the ONP concept to voice telephony services".

[4] Commission of the European Communities, Legal Advisory Board (1989): "A legal framework for Videotex/Audiotex services".

[5] Legal Advisory Board (1992): "Legal questions of European Audiotex and Videotex".

[6] British Telecommunications Engineering, July 1990, vol. 9, p. 88-97, Walters, R. E., Rose, K. R.: "Voice processing systems in British Telecom".

[7] ETR 096 (1993): "Human factors guidelines for the design of minimum phone based user interface to computer services".

[8] ISO/IEC JTC 1/SC 18 N4227, 29 October 1993: "Information technology - Document processing and related communications - User interface to telephone-based services: voice messaging applications".

[9] CCITT Recommendations Q.400-Q.490: "Specification of Signalling Systems R2" and CCITT Recommendations Q.310-Q.490: "Specification of Signalling Systems R1 and R2".

[10] CCITT Recommendations Q.700-Q.716 (1988): "Specifications of Signalling System No. 7".

[11] CCITT Recommendations Q.721-Q.766 (1988): "Specifications of Signalling System No. 7".

[12] CCITT Recommendations Q.771-Q.795 (1988): "Specifications of Signalling System No. 7".

[13] AMIS Group, Draft version 1, issue 2, (February 1992): "Audio Messaging Interchange Specification (AMIS) - Analog protocol".

[14] AMIS Group, Version 1, issue 2, (February 1992): "Audio Messaging Interchange Specification (AMIS) - Digital protocol".

[15] CCITT Recommendation X.440 (1992): "Message handling systems: Voice messaging system".

[16] CCITT. Data communication networks - Message handling systems - Recommendations X.400-X.420. Geneva, 1989. ISBN 92-61-03721-6.

[17] CCITT Recommendation G.711 (1988): "Pulse code modulation (PCM) of voice frequencies".

[18] CCITT Recommendation G.722 (1988): "7 kHz audio-coding within 64 kbit/s".

[19] CCITT Recommendation G.726 (1991): "40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM)".

[20] CCITT Recommendation G.721 (1988): "32 kbit/s adaptive differential pulse code modulation (ADPCM)".

[21] CCITT Recommendation G.728 (1992): "Coding of speech at 16 kbit/s using low-delay code excited linear prediction".

[22] CCITT Recommendation V.42 bis (1990): "Data compression procedures for data circuit terminating equipment (DCE) using error correcting procedures".

[23] ITU-T Recommendation X.25 (1993): "Interface between data terminal equipment (DTE) and data circuit-terminating equipment (DCE) for terminals operating in the packet mode and connected to public data networks by dedicated circuit".

[24] CCITT Recommendation V.32 bis (1991): "A duplex modem operating at data signalling rates of up to 14,400 bit/s for use on the general switched telephone network and on leased point-to-point 2-wire telephone-type circuits".

[25] ITU-T Q.1200 series of Recommendations (1993): "Q-series intelligent network Recommendations structure".

[26] CCITT F.400-F.422, F.500 series of Recommendations (1988): "Message handling and directory services - Operations and definition of service".

[27] CCITT Recommendation F.440 (1992): "Message handling services: The voice messaging service".

[28] CCITT Recommendations X.500-X.521 (1988): "Data communication networks directory".

[29] ISO 7498-2 (1989): "Information processing systems - Open systems interconnection - Basic reference model - Part 2: security architecture".

[30] The APS Alliance (November 1993): "Asynchronous Protocol Specification APS".

[31] ETS 300 143 (1994): "Integrated Services Digital Network (ISDN); Audiovisual services, Inband signalling procedures for audiovisual terminals using digital channels up to 2 048 kbit/s".

[32] ETS 300 144 (1994): "Integrated Services Digital Network (ISDN); Audiovisual services, Frame structure for a 64 kbit/s to 1 920 kbit/s channel and associated syntax for inband signalling".

[33] ETR 095 (1993) : "Human Factors (HF); Guide for usability evaluations of telecommunications systems and services".

[34] ISO 9241-1: "Ergonomic requirements for office work with visual display terminals (VDTs)".

[35] CCITT Recommendation E.161 (1988): "Arrangement of figures, letters and symbols on telephones and other devices that can be used for gaining access to a telephone network".

[36] France Telecom (1991): "Recommendations aux partenaires des Service Vocaux Téléphoniques".

[37] ITU-T Recommendation P.85 (1994): "A method for subjective performance assessment of the quality of speech voice output devices".

[38] ISO 639 (1998): "Code for the representation of names of languages".

[39] Voice Messaging User Interface Forum (VMUIF) (1990): "Specification document".

[40] ITU-T Recommendation P.80 (1993): "Methods for subjective determination of transmission quality".

[41] ETR 116 (1994): "Human Factors (HF); Human factors guidelines for ISDN terminal equipment design".

[42] ETS 300 001 (1994): "Attachments to the public switched telephone network (PSTN): General technical requirements for equipment connected to an analogue subscriber interface in the PSTN (Candidate NET 4)".

[43] ETS 300 114 (1991): "Attachments to the public switched telephone network (PSTN): Basic attachment requirements for modems standardised for use on the PSTN (Candidate NET 20)".

3 Definitions, symbols and abbreviations

3.1 Definitions

For the purposes of this ETR, the following definitions apply:

Access service: a service offered on first access to a network or a distant Private Automatic Branch eXchange (PABX) which permits the caller to reach the required recipient or service.

Call answering: the automatic answering of a telephone call by presentation of a pre-recorded voice announcement or greeting.

Database access: a service which permits access to a central database for the purpose of information entry or retrieval.

Interactive service: a service that permits users to select options from a menu of programmed choices in order to control the course of the session.

Intermedia: the conversion of the data representation from one medium to another.

Messaging: person to person communication that is non-simultaneous. Messaging may be "one-way", such as in telephone answering, where the caller can only leave a message, or "two-way" messaging, where the communicators can both send and receive messages.

Multimedia: the use of several media in a same session.

Stored voice service: a telecommunication service that involves the use of stored voice.

Teleaction: remote control by use of a telephone.

Telematic services: international telecommunication services, excluding telephone, telegraph and data transmission services, offered by Administrations and defined by the CCITT for the purpose of exchange of information via telecommunications networks. Examples of telematic services are the telex service, the facsimile service, Message Handling Services (MHS) and Videotex service. Telematic services provided on an ISDN may be considered as teleservices.

Teleservice: a type of telecommunication service that provides the complete capability, including terminal equipment functions, for communication between users according to protocols established by agreement between Administrations and/or Recognised Public Operating Agencies (RPOAs).

Transaction: an operation which is regarded as a two way interactive activity using read/write access to a service.

Videotex service: an interactive service which provides, through appropriate access by standardized procedures, for users of Videotex terminals to communicate with databases via telecommunication networks.

Voice bulletin board: a facility which enables messages to be "posted" in a public or restricted access mailbox, so that any callers can access them. The bulletin board differs from voice mail in that Voice Mail delivers messages to specific individuals or groups of individuals.

Voice Mail: voice messages that are created by accessing a voice messaging system for delivery to one or more subscribers. Premeditated, non-real time communication using Voice Mail system technology.

Voice message delivery: the facility of a voice messaging system to call a phone number for delivering a voice message to a recipient. This may be to a Voice Mail subscriber for delivery of new or urgent messages, or to a non-subscriber who is receiving a voice message addressed with a telephone number. Delivery may be contingent on password entry.

Voice messaging: a general term which includes Voice Mail and telephone answering. See Voice Mail.

Voice storage: a facility to memorize voice information.

3.2 Symbols

For the purposes of this ETR, the following symbol applies:

kbit/s	Kilo-bit per second

3.3 Abbreviations

For the purposes of this ETR, the following abbreviations apply:

ADPCM	Adaptive Differential Pulse Code modulation
AMIS	Audio Messaging Interchange Specification
API	Application Programming Interface
APS	Asynchronous Protocol Specification
ASCII	American Standard Code for Information Interchange
ASR	Automatic Speech Recognition
AU	Access Unit
B-ISDN	Broadband ISDN
CEC	Commission of the European Communities
CELP	Coded Excited Linear Prediction
CLI	Calling Line Identity
CNET	"Centre National des Etudes de Télécommunications"
CNIL	"Conseil National Informatique et Liberté" (in France)
DDI	Direct Dialling In
DID	Direct Inward Dialling
DPNSS	Digital Private Network Signalling System
DRG	"Direction de la Réglementation Générale" (in France)
DSP	"Direction du Service Public" (in France)
DTMF	Dual Tone Multi-Frequency
E-mail	Electronic Mail
EHDV	"Ergonomie et Harmonisation des Dialogues Vocaux"
EPS	Electronic Publishing Services (Publications) Ltd.
Fax	Facsimile
FCC	Federal Communications Commission (in USA)
FTAM	File Transfer Access & Management
G3	Group 3 (analogue facsimile)
G4	Group 4 (digital facsimile)
GSM	Global System Mobile
ICSTIS	Independent Committee for the Supervision of Standards of Telephone Information Services
IEC	International Electrotechnical Commission
IN	Intelligent Network
IPAV	"Interface Programmable pour Application Vocale"
ISDN	Integrated Services Digital Network
ISO	International Standards Organisation
ITU-T	International Telecommunications Union for Telecommunications (former CCITT)
JTC 1	Joint Technical Committee one (ISO/IEC)
LAN	Local Area Network
LAPB	Link Access Protocol-Balanced
LD-CELP	Low-Delay Code Excited Linear Prediction
MAN	Metropolitan Area Network
MCN	Mobile Cellular Network
MFC-R2	Multi-Frequency Code (Region 2)
MFPB	Multi-Frequency Push Button
MHS	Message Handling System

MMI	Man Machine Interface
MS	Message Store
MTA	Message Transfer Agent
MTS	Message Transfer System
MVP	Multi-Application Voice Processing
MVIP	Multi-Vendor Integration Protocol
NERA	National Economic Research Associates
N-ISDN	Narrow-band ISDN
NW	Network
ONP	Open Network Provision
OSI	Open System Interconnection
PABX	Private Automatic Branch eXchange
PAD	Packet Assembly/Disassembly
PBI	Phone Base Interface
PC	Personal Computer
PCM	Pulse Code Modulation
PSDN	Packet Switched Data Network
PSOLA	Pitch Synchronous OverLap and Add
PSPDN	Packet Switched Public Data Network
PSTN	Public Switched Telephone Network
PTN	Public Telephone Network
PTT	Postal, Telephone, and Telegraph
RPE-LTP	Residual Excited Linear predictive
RPOA	Recognised Public Operating Agency
SCSA	Signal Computing System Architecture
SS7	Signalling System Number Seven
SVS	Stored Voice Server
TCP/IP	Transmission Control Protocol/Internet Protocol
TARM	Telephone Answering and Recording Machine
TE	Terminal Equipment
UA	User Agent
UK	United Kingdom (of Great Britain and Northern Ireland)
UMIG	Universal Messaging Interoperability Group
USA	United States of America
VMAE	Voice Mail Association of Europe
VMUIF	Voice Messaging User Interface Forum
VSELP	Vector-Sum Excited Linear Prediction
WAN	Wide Area Network

4 Audiotex (stored voice services)

4.1 Definition

The terms of reference required the Project Team "to define an Audiotex service and its relation to other services (e.g. Teleaction)".

Initial studies revealed significant confusion in the use of the term Audiotex, with interpretations varying from country to country and from user to user.

The original CEC mandate [1] stated that "Audiotex services are telematic services, which use standard telephone equipment to convey information to users. They can be either interactive or non-interactive". This suggests a fairly all embracing definition.

The Electronic Publishing Services (EPS) report on the Audiotex information services market in Europe [2] restricted its definition to interactive access to databases and did not include non-interactive databases or voice messaging.

The National Economic Research Associates (NERA) study of the application of the ONP concept to voice telephony services [3] included both active and passive interrogation of databases and even included live broadcasts and conference services within its definition of Audiotex. Voice Mail, however, which was defined as including voice storage and retrieval and telephone answering services, was not included, and was treated as a separate service.

The Legal Advisory Board of the CEC in 1989 [4], described Audiotex as "a means of providing requested information via an auditive form" (analogous to Videotex). Later, in 1992 (see CEC Legal Advisory Board, "Legal questions of European Audiotex and Videotex") [5], it defined Audiotex as "an interactive service which provides access via the telephone network to databases, some of which use voice recognition or sound recognition. Information is provided in voice form or by means of other media such as telecopies and Videotex".

Other authors (Walters and Rose in British Telecommunications Engineering [6]), writing on voice processing systems in general, considered Audiotex as one of a family of such systems, restricted to database interrogation, and described it as "the juke box of the voice world".

In France, (P. Devauchelle, private communication) the definitions most often found cover the whole range of services accessible over a telephone that return information in a vocal form. The confusion is greater because a reference to "Audiotex" often carries an implicit reference to "Videotex", which has a large field of application.

ETR 096 [7] describes Audiotex as either a generic term for all services, or simply inquiry access services.

It can be seen that there are many differing views as to what constitutes an Audiotex service and, in formal usage (e.g. in lectures or articles on the subject), a distinction is always drawn between voice messaging and the other services (information, consultation, transaction, etc.).

It is thus misleading to use the word Audiotex without a previous definition. The opinion given in this ETR is that it is better not to use it at all in view of the confusion surrounding its use. It is clear that all of the services being reviewed are voice services in the broadest sense although they may not all be network services. Furthermore, all of the services utilise some sort of voice storage.

Therefore, the term "s**tored voice services**" is used in this ETR to describe the whole body of such services, and each separate application or service is distinguished by a more precise definition (e.g. information, consultation, transaction, etc.).

Therefore, to avoid confusion in the description and specification of stored voice services, it is recommended that:

Recommendation 1

The use of the term stored voice services should be encouraged and the term Audiotex be deprecated in the field of standardization.

4.2 Classification

Work at present going on in International Standards Organisation/International Electrotechnical Commission committee (ISO/IEC JTC1/SC 18) [8] on voice messaging applications, contains a taxonomy of telephony based services which has been drawn on to provide the classification given in figure 1 (see also ISO/IEC JTC1/SC18 N4227 [8]).

In this classification, stored voice services are shown as a subset of general telephone based services. Services such as transaction and messaging may be realised in more than one medium, for example, messaging may be via a text service or via facsimile as well as provided by a voice service.

Access services are interpreted as those offered on first access to a network or to a distant PABX.

Messaging services are those where a message is left in a store for later delivery.

Transaction is regarded as a two way interactive activity using read/write access to a service.

This classification contains a class for database entry as well as the class for database retrieval shown in the ISO/IEC work.

Database information entry is classed as the entry of information to a central database where the message is interpreted by the database as a code rather than a voice message.

Database information retrieval is classed as the read only access to stored voice information.

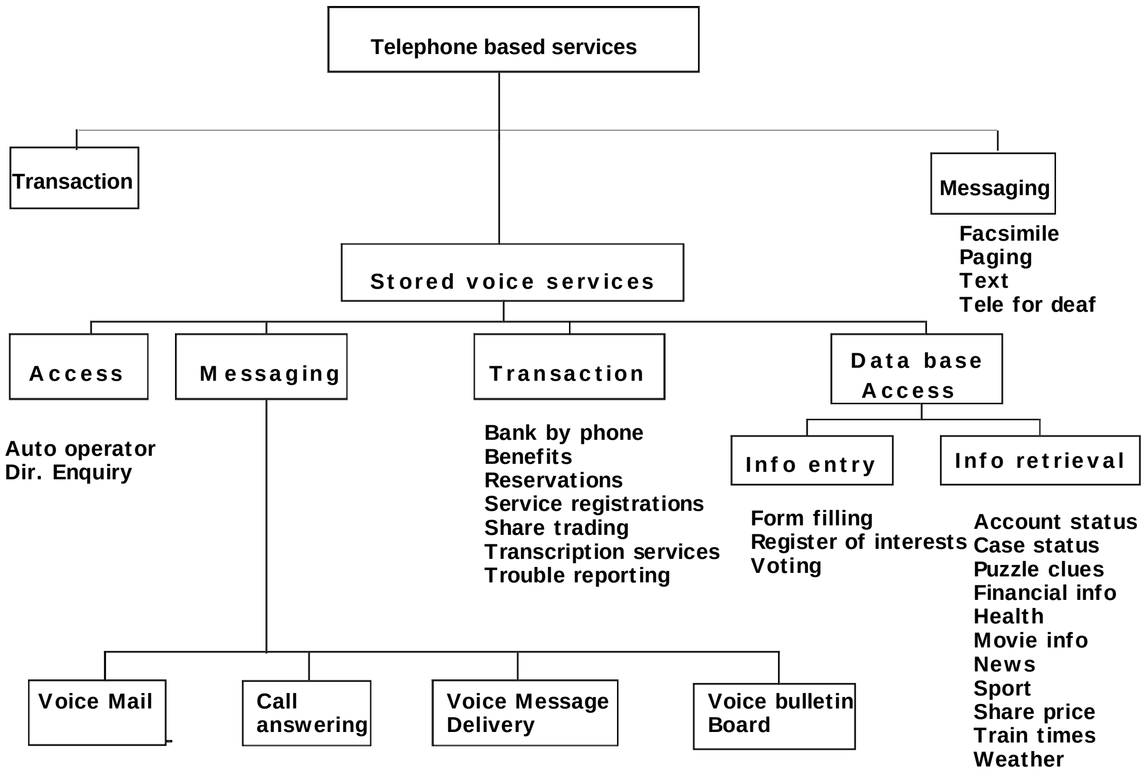


Figure 1: Classification of stored voice services

5 Questionnaire

5.1 Analysis of replies

Of **480 questionnaires** addressed to the ETSI members and non members, only **31 replies** were received (about 7%).

The replies have been sorted into two main classes:

- manufacturers, developers, R&D labs, advisers: **23** replies**;**
- **s**ervice providers and network operators: **8** replies**.**

The telephone answering application seems to be the basic kind of stored voice service application. Answering, together with Voice Mail and database interrogation are the most common applications offered by the manufacturers. Voice Mail and auto-attendant seem to be the most common among the services offered by the network operators and service providers but the evidence is not very conclusive due to the lack of replies.

Database input and transactions appear to be offered by many manufactures but are not much referred to by the network operators and service providers. Numerous other applications are offered by the manufacturers but in fact do not, so far, appear to be used very much.

Among those who have replied few participate in the standardization process in official organisations such as e.g. ISO and ITU-T, but there is a significant activity in the user and manufacturer associations.

The main associations referred to in these replies are:

- EHDV (France): "Ergonomie et Harmonisation des Dialogues Vocaux";
- IPAV (France): "Interface Programmable pour Applications Vocales";
- MVIP (USA): "Multi-Vendor Integration Protocol" (for the voice-bus interface);
- SCSA (France): "Signal Computing System Architecture" (for the voice-bus interface);
- VMAE (Switzerland): "Voice Mail Association of Europe";
- VMUIF (USA): "Voice Messaging User Interface Forum".

There is a reference to "CNET", which seems to be a reference to the EHDV association.

18 various definitions and explanations of the meaning of "Audiotex" have been noted, which confirms that the use of this term is confusing and misleading.

It is clear that the main terminal used is the normal telephone, with the stored voice services invoked by the use of the dialling push button. Decadic signalling is offered by most manufacturers but referred to by only 50 % the service providers. Most manufacturers offer the use of ISDN and Mobile Telephone sets to invoke Voice Database as well, but once again they are only mentioned by 50 % the service providers.

By using voice activation (in the speech band) it is clear that any telephone set can be used to invoke a Voice Database including voice recognition facilities. Voice recognition, although being a feature offered by some manufacturers, is not a mature technology at present because for the time being, multi-speaker recognition is not very reliable and offers only a limited vocabulary. Some very simple voice activation techniques using the difference between noise and silence (grunt recognition) as a command are also used for simple systems.

One company referred to activation through other channels, such as GSM digital signalling, and Videotex.

The voice servers are either stand-alone systems connected to network access interfaces, or included within networks. Some special accesses still exist but the trend is to use standard interfaces. Manufacturers are beginning to offer voice servers to be used as part of Intelligent Networks.

Special network interfaces are also used, especially Local Area Networks (LANs) such as Ethernet, but it is not clear whether the network interface is used to carry the voice data or to interconnect the server to some external databases. The interconnection through Wide Area Network (WAN) is not widely referred to.

The most popular inter-register and common channel signalling systems are both DTMF and Multi-Frequency Code (Region 2) (MFC-R2), in both analogue and digital networks. Nevertheless there is an increasing trend to offer ISDN primary interfaces. Signalling System No.7 (SS7) [10], [11], [12] is also used, probably in case where the server is included in the network. In some private installations, the Digital Private Network Signalling System (DPNSS) is also used.

Most network interfaces offer Direct Dialling In (DDI) or Direct Inward Dialling (DID) facilities to access different services in the same server.

The main interchange protocols used today are the Audio Messaging Interchange Specification (AMIS) analogue [13] and digital protocols [14] developed by a manufacturer/user association in America. The other protocol mentioned is CCITT Recommendation X.440 [15] which is the application level adaptation of CCITT Recommendation X.400 to X.420 series of recommendations [16] for the specific case of voice-data transfer. CCITT Recommendation X.440 [15] could be used as a part of the AMIS interchange protocol at the application level.

The main services which are linked to voice services are paging and group 3 facsimile. The paging service is one of those which is the most used. The combination of voice services with Videotex is proposed by some manufacturer, although this kind of interaction is not commonly used for the time being. The few answers referring to interaction with telex confirms the obsolescence of the latter. The interaction with group 4 facsimile seems to be not yet available. There is also interaction with a number of other services such as E-Mail, telephone services, Intelligent Network services and GSM.

Most manufacturers support the Premium Rate Billing system and most service providers are using it, but some other billing systems are used as well and very often the service providers are using both.

The voice services as for the other telematic services like Videotex are subject to regulation in some countries. The main regulatory bodies mentioned are the Independent Committee for the Supervision of Standards of Telephone Information Services (ICSTIS) in the United Kingdom (UK), the "Conseil National Informatique et Liberté" (CNIL) and the "Direction du Service Public" (DSP) in France, the Federal Communications Commission (FCC) in the United States of America (USA) and, in some countries where telecommunications are not yet deregulated, the Postal, Telephone and Telegraph organisation (PTT) themselves. In some countries it seems that the common law is sufficient. Sometimes the type approval authority is mentioned by reference to the "Direction de la Réglementation Générale" (DRG), which probably refers to the access requirements rather than the regulation or the code of practice dealing with the use of voice services.

Only a third of the replies said that there are some applicable laws, but the rest were left blank, because the actors are probably not aware of them. It is also possible that the recipient who has replied to the questionnaire was not skilled in legal aspects.

User guidance is generally given both in a user guide and in an abbreviated form, but the on-line menu, although proposed by some manufacturers, is not identified as used significantly by network operators and service providers.

For the control of the voice service by the user, the use of voice and of numeric keys are both described, but the use of the numeric keys is more common than voice, and in most cases the numeric keys are used with, in addition, the * and # keys.

The question "what kind of dialogue is used" has produced misleading choices which cannot lead to a clear conclusion. Nevertheless, it seems that interactive dialogue is as commonly used as are key and voice command.

5.2 Comment

It is difficult to draw firm conclusions from the answers received to the questionnaire because of the very poor response. Nevertheless, it is possible to make some comments based on those answers that have been received and supported by previous knowledge.

Although database input and transactions are offered by some manufacturers they are not yet widely used, possibly because of their complexity. This may be due to the difficulty of operation by voice prompting, together with the low level of the security and the lack of reliable recognition of spoken input numbers.

The use of auto-attendant services seems to be increasing, particularly in areas where DDI is not available. This growth is in spite of user dissatisfaction.

It is clear that the main terminal used is the normal telephone, with stored voice services invoked by the use of the push button dialler. Although the decadic end-to-end signalling is not reliable, it was in the beginning the only available means of operation, but the use of this technique is now decreasing with the wider spread of DTMF signalling in the European countries. Therefore, the Multi-Frequency Push Button (MFPB) telephone set (DTMF) will become the main end to end signalling system to invoke a Voice Database. ISDN and mobile telephone sets are more likely to be used to invoke voice database as well, the use of the mobile telephone increasing more rapidly than the ISDN one.

ISDN terminals and mobile telephones have to offer DTMF signalling features at present in order to allow the access to any existing voice database, although these terminals are able to interact with their own voice databases or their own short message systems (offering a message on their visual display) through their own signalling channels.

In conclusion, for the time being, the most popular signalling systems remains DTMF and voice activation. The multi-speaker voice recognition technique, although not mature yet, seems to be the technique of the future which will increase the potential of stored voice services.

6 The use of stored voice services

6.1 Existing services

There are a large number of stored voice services at present in use and, as the field is growing rapidly, it is difficult to give an exhaustive list. The following is a set of the most important examples based upon the classification given in figure 1 in subclause 4.2.

6.1.1 Access services

Auto attendant

The function of an auto attendant service is to imitate the operation of a human switchboard operator. It is usually placed at the primary answering point and answers incoming calls, giving an announcement to the caller, and inviting him to input the extension number of the person he wishes to reach. Such a usage is sometimes referred to as "poor man's Direct Dialling In (DDI). Most implementations of auto attendant offer more complex interactions such as directory searching and reversion to a live operator to ensure that all calls are successfully terminated.

Automatic salutation

Auto salutation is a form of auto answer which enables a PBX operator to record a personal welcoming message which is then offered automatically from a voice store whenever the answer key is operated.

Directory enquiry

Directory enquiry services are normally an adjunct to a human operator whereby, when the required destination is established, the number can be transmitted to the caller from a voice store whilst the operator deals with the next enquiry.

6.1.2 Messaging

Voice messaging can offer a number of services, such as:

Call answering

Call answering is the simplest function of a voice messaging system whereby a call is answered and the caller is invited to leave a message. This function is well known when provided by a simple Telephone Answering and Recording Machine (TARM) but can also be provided by a voice messaging system accessed by a call diversion service so that it acts as a shared TARM. When the system is integrated with a public or private network so that information on the original destination of the call is available to the answering system, more complex message options can be provided. This type of service is particularly valuable on mobile networks where the called party may often be switched off or out of range. Privately offered call answering systems find difficulty in competing with those offered by network operators due to the tariff structure.

Voice mail

A voice mail system is the voice equivalent of electronic mail (E-mail) which provides a number of personal mail boxes, each capable of containing a number of voice messages. A typical voice mail system will offer the ability to reply to a voice message without needing to know the address of the originator and the ability to attach a received message to another voice message and then forward them together to a third party.

Voice message delivery

Voice message delivery allows a user to record a message for automatic delivery to a specified addressee, often at a later specified time. This facility can also be used as a broadcast facility, allowing a message to be sent simultaneously to several addressees.

Voice bulletin board

Voice bulletin boards are analogous to E-mail bulletin boards, whereby, after inputting a security code, the caller is allowed to interrogate and leave messages on one or more bulletin board. Systems without security codes are often used on premium rate services e.g., "contact magazines" or for "messagerie roses".

It is common, particularly in large private networks, for a number of voice messaging systems to be able to be accessed, one from another, so that messages can be sent across the network or via a public network from one network to another. Such a facility requires the ability for automatic signalling across networks, and the use of standard protocols.

Some systems have gateways to other messaging systems such as paging systems, so as to allow an addressee to be informed that a message is waiting.

Many modern voice messaging systems can also handle facsimile messages, accepting an incoming facsimile into its voice store, treating the modem signals as voice signals, and then offering the stored facsimile message to the addressee, normally with a recorded voice message requesting to which number the stored facsimile should be delivered. The original modem signals are then reproduced and sent to the facsimile machine specified.

6.1.3 Transaction

A large number of transaction systems are already in use and the range of services offered is continually growing. Some services are public, offering open access to a free service or one utilising a premium rate (kiosk service). Most are for closed user groups, requiring the input of a personal access code, some being free and others by subscription.

It is sometimes difficult to differentiate between a transaction and an interactive access to a database, especially when information entry is involved, but in general, a transaction tends to be more interactive, with the system responding differently dependent on the answer, and finally giving some sort of active output. Database entry tends to be interpreted as a more passive process of the input of a set of answers to a fixed set of questions for the purposes of information collection.

Typical examples of transaction services are:

Banking by phone
Banking systems offered over the phone offer a range of facilities ranging from a simple interrogation of an account balance to full bill payment facilities. Many systems provide voice recognition and accept voice input (with a rather restricted vocabulary) as well as input from a telephone keypad.

Reservations
Allow the interrogation of a service, for example a flight timetable, the booking of a reservation, and possibly the payment by means of a credit card.

Route finding
A service that will describe the best route between two towns (identified by their postcodes). The answer can be provided by voice message or by facsimile.

Share trading
Share trading permits the purchase and selling of stocks and shares over the telephone by the member of a closed user group generally with settlement by means of a suitable credit account.

Trouble reporting
Allows a customer to report a problem with an item of equipment and to make suitable arrangements for its repair.

6.1.4 Database entry

Form filling
Most database entry is, in effect, a sort of form filling, in that the caller generally provides answers in response to a set of stored questions. One interesting example of this usage has been in the Netherlands, where the EC Directive to establish a nation-wide registration system for cattle has been implemented through use of data entry on a stored voice system.

Register of interests
This is a form of database entry which allows a caller to register his interest in a range of subjects in order subsequently to be placed for example on an appropriate mailing list.

Voting

This is a particular form of data entry where the entries made after listening to a recorded message are counted to form a vote, for example to decide the winner in a television talent contest.

6.1.5 Database information retrieval

Database information retrieval covers a range of services ranging from simple passive non-interactive systems, where a service is contacted at a single address, through more complex systems, where different messages are accessed by (typically) different last digits, to fully interactive systems where different messages are received in response to differing signals sent to the service after answer.

The services offered can appear under various names and offer a vast range of information, as varied as can exist in the mind of any enterprising service provider.

Speaking clock

One of the simplest database information retrieval services is the "speaking clock" which is the oldest of the stored voice services. It is a typical example of a non-interactive stored voice service, with the message constructed from concatenated phrases. It is usually provided at a normal rate on a short, national telephone number address.

Cinema programmes

This is an example of a slightly more complex service with interaction which allows interrogation of the programme for a particular cinema in a particular town.

Horoscope

This service provides a horoscope either by different telephone numbers representing star signs or by an interactive process whereby a birth date is entered during the call. More complex services are similar to a transaction providing a written output by facsimile.

Credit rating

A service that gives information on the financial status of companies. Usually provided to a closed user group and working on an interactive basis. Some systems use voice recognition.

Charity fund-raising

Funds are sometimes raised by means of a message describing the charity provided on a premium rate service.

Competitions

These are interactive services where the caller can win prizes by giving the correct responses to recorded voice messages. These competitions are run on premium rate services and are often fraudulent in that the answers are very simple, but the call charges outweigh the value of any available prize.

6.2 Future services

Future services will grow as fast as entrepreneurs can think of new ways of profiting from stored voice services. Multimedia services will become more common, and voice responses will be provided to guide users through teleaction procedures such as the remote control of central heating or home security.

Intelligent terminals on intelligent networks will be able to offer more user-friendly vocal help information. The use of voice recognition will grow as the algorithms improve and recognition becomes more reliable. It can be expected that in a few years, arising from present on-going research, automatic language translation services will be offered. In the future, such a service can be expected to be available in real time.

Future developments will be particularly affected by new terminal facilities offered by the ISDN, and by the greater use of more complex terminals such as multimedia terminals. It can be expected that the display facilities available on multimedia terminals will tend to replace some of the provision of information in stored voice form. On the other hand, stored voice messaging may well be considered as a useful adjunct to displays, and can give useful assistance to the visually handicapped.

It can be expected that more complex services will be offered, generating more links between the various kinds of messaging services as the usage of non-telephony services grows.

7 Technologies used

7.1 General

The increasing convergence of data processing and telecommunication techniques has caused many changes in recent years. The rapid technological development of digital switches has added many new potential services to the basic facility of voice telephony. The innovative element in voice services is based on the fact that the information is conveyed by means of computers. Thanks to the digitization, coding and compression of the voice, the data can easily be stored and conveyed as alphanumerical data either through digital networks in digital form or through analogue networks in analogue/digital and digital/analogue coding and decoding form by using modems.

Stored voice services are mainly supported by techniques which involve spoken information conveyed over the telephone network, mediated by machine slave control which permits the following features:

- **interworking** between user and machine under the user control;

- **communication** between users mediated by the machine on real time or with time delay;

- **transaction** carried out by command executed in a third party information system.

In addition, some functions deal with the marketing aspects in order to allow the various charging options, mainly:

- the **freephone service** where the charges are paid by the provider of service;

- the **dial-up service** where the caller pays the communication cost only, the information service being free;

- the **premium rate service** or **kiosk service** where the caller not only pays the communication cost but also remunerates the service provider.

Access to stored voice services should be possible from any telephone set (basic telephone set, ISDN terminal, mobile phone, etc.) and at the minimum by the basic standard telephone set requiring no additional specific function so as to give wide availability.

The central element supporting the stored voice services is basically a sophisticated machine based on computer techniques which is generally designated as the "voice server", or simply the "server", which may be designed to operate on various type of platforms.

It is difficult to make an exhaustive description of a voice messaging platform, because its architecture and its features are varied and dependant upon market needs and the networks which they access.

A voice platform, whether it is a PBX, Personal-computer, mini-computer, mainframe computer, or specific machine, needs to be accessible simultaneously by a significant number of users. Therefore, the network interfaces, recording and generation features for the voice messages, voice storage feature and media, user command interpreter, external interface to the database and application software all need to be sized accordingly.

There is also a need to interconnect a server to another server in order to interchange the voice messages, or to interchange the voice data between a server and a recipient machine or a private network without human intervention. Protocols allowing these end-to-end interchanges between, and through, heterogeneous environments need to be considered.

There is a great market demand for interaction with other services (as, for example, facsimile transmission or paging) to provide additional features, therefore access through some translation systems to such services or to other kind of terminal equipments needs to be provided.

Taking into account the above considerations, the following main components are required to support the stored voice services:

a) terminal equipment to access the voice services;

b) voice servers with computer control by, noise/silence, tones or voice interpreter, with voice or tones generator and with memory storage;

c) network interfaces;

d) external interfaces to databases;

e) interchange protocols.

7.2 Terminal equipment used to access voice services

Voice services may be accessed from any item of speech terminal equipment connected to the network, either directly or through a private network, such as:

- a basic telephone with MFPB;
- any telephone (voice recognition case);
- an ISDN terminal (with DTMF signalling feature);
- a mobile terminal (with DTMF signalling feature);
- a PABX extension Terminal;
- a key-system terminal.

7.3 Voice server

7.3.1 General

The central equipment of the voice service, hereinafter called the server, is basically a complex machine based on computer techniques, accessible by the use a of simple telephone and allowing some intercommunication facilities with other voice servers or other complex systems.

The server is designed on different type of platforms which are mainly the following:

1) PBX-based application platform;

2) stand-alone application platform;

3) network-integrated or network-peripheral application platform;

4) PC-based application platform.

To begin with (early 1990s) there were mainly two families of voice processing application: voice mail and (interactive) voice response. Now Multi-Application Voice Processing (MAP) is emerging allowing in one call session the use of several voice processing applications supported by the voice service. The trend is to introduce the concept of a "universal mailbox" in which voice, facsimile, and E-mail messages can be stored and manipulated in response to user command.

Although it is difficult to describe such a complex system it is possible to identify the following main physical and logical elements:

Network interfaces:

Analogue or digital interfaces provide for incoming and outgoing calls, tone detection, dialling, etc., with DDI facilities in most cases. The servers access the networks directly or through PBXs. These interfaces are generally the standardized access to the network. Sometimes when the server is more and less integrated in a network, they are specifically interconnected to the heart of the network on the standardized signalling system such as the Signalling System No.7 (SS7), the CCITT Q.700 to Q.716, Q.721 to Q.766 and Q.771 to Q.795 series of recommendations [10], [11], [12] or will be interconnected to the future world-wide signalling system (Q-SIG). They can be also interconnected to some vendor-independent standard signalling system, such as the Digital Private Network Signalling System (DPNSS), and to some proprietary standards, such as Cornet of Siemens, ABC of Alcatel.

Voice recording:

This function digitizes and codes the analogue audio signal, and records and in some case transcodes the digital signal. Compression techniques are often used to reduce the size of the data file to be stored.

Voice output:

This function converts the digital voice data in real time into analogue voice (PSTN) or into a structured digital transmission at a rate compatible with the network used (ISDN, PCM, GSM, etc.). Some other techniques for speech synthesis can be also used such as Text to voice conversion (e.g. the France Telecom PSOLA technique).

Voice storage:

The voice can be stored in analogue form on a magnetic tape medium for simple passive voice servers or answering machines but this technique is rapidly becoming obsolete and now most storage is now performed in digital form on mass storage media such as a magnetic hard disk, tape or optical disk memories. The voice is generally stored in a voice mailbox or in a universal mailbox.

Voice interchange:

In order to allow the interchange of voice messages between server and complex terminal equipment there is a need for a set of interchange protocols such as e.g. AMIS protocols ([13] and [14]). For the analogue interchange there exist some specific protocols which are likely to become obsolete in the future, although some manufacturers need to implement them in their systems for economic reasons. Probably a digital interchange system according to the OSI model will become imperative in the near future in order to allow interchanges through heterogeneous environments. Bearing in mind that most of standards exist for the layers up to the transport data function level and for some layers of the distributed data processing (except maybe for the application level), some functional standards or profiles standards need to be used.

User command interpreter:

This feature interprets the commands sent by the user from the telephone. The commands are mainly in the form of DTMF, single tones, silence/noise or speech.

7.3.2 Speech coding and compression

In a voice server, the voice and sound are digitized and coded in various medium and low bit-rate speech coding systems which are required to fulfil several conflicting objectives. These are mainly to achieve world-wide connectivity, to achieve optimum quality on an end-to-end basis and to reduce the size of the digital record in term of number of kilo-bytes for a given speech length.

The reduction in stored data serves to reduce the record space in the storage memory and to reduce the amount of data to be transmitted through an interchange process for a given message. This latter aspect is significant when transmitting the voice data on the PSTN by means of modems.

In some case, as is the case for radio digital transmission, the speech coding rate is chosen to be as low as possible in order the make an effective use of the available radio frequency channel space by reducing the bandwidth required.

The basic speech coding system used is Pulse Code Modulation (PCM) coding at 64 kbit/s according CCITT Recommendation G.711 [17] (bandwidth 3,1 kHz), and at 56/64 kbit/s according the CCITT Recommendation G.722 [18] (bandwidth 7 kHz) which give a good quality (1 to 0 qdu) but are not optimised in terms of rate. It is the basic rate for the PCM and ISDN B-channel transmissions.

The other main coding systems are, the Adaptive Differention Pulse Code Modulation (ADPCM) coding at 32 kbit/s according to CCITT Recommendation G.726 [19] (which replaces CCITT Recommendation G.721 [20]) and the Low-Delay Code Excited Linear Prediction (LD-CELP) coding at 16 kbit/s CCITT Recommendation G.728 [21], which gives similar quality (3,5 qdu) (but lower than PCM coding).

The coding system used for GSM is the Residual Excited Linear predictive (RPE-LTP) coding at 13 kbit/s which gives a 7 qdu to 9 qdu quality.

Other coding techniques are sometimes used such as the Vector-Sum Exited Linear Prediction (VSELP) coding at 7,9 kbit/s (IS-54) developed by the ANSI Telecommunication Industry Association in the USA, and some other coding techniques are in the standardization process in USA and the Code Excited Linear Prediction (CELP) at 4,8 kbit/s. CCITT Recommendation G.726 [19] also permits other bit rates such as 40, 24 and 16 kbit/s which give lower quality.

GSM is also in the process of selecting another process at about 6 kbit/s to 7 kbit/s rate for the next generation of digital mobile radio.

To carry the digital voice data or other data by an interchange process through the analogue network (or leased line) by means of modems there are also some compression techniques (CCITT Recommendation V42 bis [22]) which compress the data before transmitting it to line and at the other end decompress the data, increasing only the transmission speed without loss of quality, the integrity of the data being supported. This does not reduce the record size. Some compression techniques aim only to reduce the storage space, keeping the data integrity as well but it takes time to compress and decompress the data, therefore increasing the response time of the system. We have also to take into account the fact that storage memory has an increasing capacity for a decreasing cost.

It should also be pointed out that to support the interaction with the user, some non-voice signals such as Dual-Tone Multi-Frequency (DTMF) need to be transmitted and for which special arrangements need to be made in those digital transmission system supporting a low-encoding rate which is not transparent to such signals (such as GSM).

7.4 Application software

The application software is in charge of controlling the various modules building up the stored voice server.

Usually, all low level processing (voice recognition, voice generation, DTMF decoding, text to voice synthesis, etc...) is performed by a set of routines distributed with the corresponding hardware, whether it is an add-on board for a personal computer, a stand-alone device or another kind of device.

Some applications are developed using an application generator. Usually, a generator is a menu-driven application, creating the application code according to the inputs of the person creating the application. These inputs include the contents of the voice prompts, the response to be performed for each user input (DTMF codes, voice input, etc...), the logical sequence of actions (user validation, user data input, data validation, output data generation, etc...). The main advantage of an application generation is the resulting short time-to-market. As far as is known, there is currently no standardization effort going on in this domain.

When flexibility is the main criterion, it is often necessary to develop the application using a lower level approach. Low-level routines in charge of handling the voice aspects of the application can be called using a defined Application Programming Interface (API). Until now, most manufacturers were delivering proprietary APIs. But it seems that a standardization process is emerging, which includes the whole telephony aspect. Several software companies, along with exchange manufacturers and computer companies have defined several telephony APIs (Windows Telephony Application Programming Interface, from Microsoft & Intel, Telephony Service, from AT&T and Novell, etc.) mainly for the personal computing world.

Other APIs are available, to implement usual interfaces (SQL API for database access, communication API for access to various networks, usually through specific hardware, facsimile generation from ASCII text, etc...). Most of these APIs are proprietary, and not specific to stored voice services.

7.5 Networks

7.5.1 General

It is necessary to distinguish the networks to which the user terminal equipment is attached, the networks to which the servers are connected and the various networks through which the information is conveyed. Furthermore, the network access can be made directly or indirectly through private complex installations such as PABXs, Centrex including other types of networks such as the PDN or a LAN.

Three kinds of voice interchange forms are to be considered:

- analogue voice in the **analogue speech** band;
- speech in **digital** form with sampling rate integrity according to specified encoding rules and rates (e.g. ISDN, B channel 64 kbit/s A law);
- **audio data** in digital form information in various bandwidths conveyed in various manners depending on the network type and on the bearer service attribute.

To convey the **digital audio data,** the voice interchange system does not require any different bearer services than for other data types. The distinction is made only at the presentation and application layers.

7.5.2 Network accessed by user Terminal Equipment

The most widespread network over which users invoke the stored voice services is the Public Telecommunication Network (PTN) with user access by means of a standard telephone directly connected to an analogue subscriber interface (PSTN access) or by digital basic access (ISDN basic access). Access to the PTN is also possible through PBXs by both analogue and digital access. Furthermore the use of voice services from cellular telephones is also growing rapidly especially through the GSM network.

There are also some other radio interfaces or radio networks from which the voice services could be significantly invoked, such as CT0, CT1, CT2-CAI, DECT, TETRA, Analogue cellular etc...

There are only two signalling systems between the user and the Stored Voice Servers (SVSs) which can be conveyed through most of the networks which support voice service. They are speech and tone signalling. Since the voice recognition currently exhibits serious reliability problems, the DTMF signalling is, for the time being, the most widespread system. As electronic exchanges spread through networks, the availability of such terminals can be expected to increase.

Where the SVS is a terminal connected on a PSTN access (as at interface 9, figure 2), the DTMF signals will suffer an extra frequency dependent line loss over and above that experienced on the line between the TE and the public exchange via interface 8. To ensure reliable operation, the DTMF receiver in such SVS terminal requires a different frequency/sensitivity skew to that of a receiver in an exchange.

A standard is required for such a receiver. Such a standard is currently on the ETSI STC-TE 5 work programme.

To permit user access to the overall SVSs, the DTMF signalling should be clearly supported on overall TE as the minimum interactive signalling service. Therefore, decadic MFPB telephone sets, mobile radio-TE, cordless radio TE, PBXs, Key-systems, ISDN-TE, PCs, and other kind of PCNs, should provide the DTMF signalling mode as available in MFPB telephone sets. Where these signals are to be sent over a digital local connection, a different frequency skew is necessary, and should be specified in the appropriate standard.

7.5.3 Network to which the server is connected

It is very common to find a stand-alone server connected to the PTN through analogue or digital interfaces. In order to allow simultaneous access of several users, voice servers, as do PBXs, use mainly primary access interfaces (at least 30 ports) such as primary ISDN access, PCM MFC-R2 or, e.g. in the UK, through 2 048 kbit/s interfaces using channel associated signalling. Nevertheless some small systems can use a number of analogue lines or basic ISDN accesses. It is also common to use ITU-T Recommendation X.25 [23] interfaces in order to convey data (or voice in form of digital data) through a PSDN network.

When the voice messaging platform is a PBX-based one, the network interfaces are provided by the PBX.

When the voice messaging platform is a part of the network, as for example a network peripheral within the concept of Intelligent Network architecture or in the radio cellular network, the signalling system is directly connected to the server through SS7 or a proprietary signalling system.

The server can be also connected to Local Area Networks (LAN) or Metropolitan Area Networks (MAN).

7.5.4 Multiple networks

It is very common to carry the data through several different networks during one communication session. It is common to pass through a private network and/or through a PBX to access the PTN network by PSTN or ISDN access as shown in figure 2.

Generally, the interaction in real time between the user TE and the SVS is made through the speech channel by voice or DTMF for example (see figure 2) as for the connections <8-4>, <7-9>, <3-5>, <11-6>.

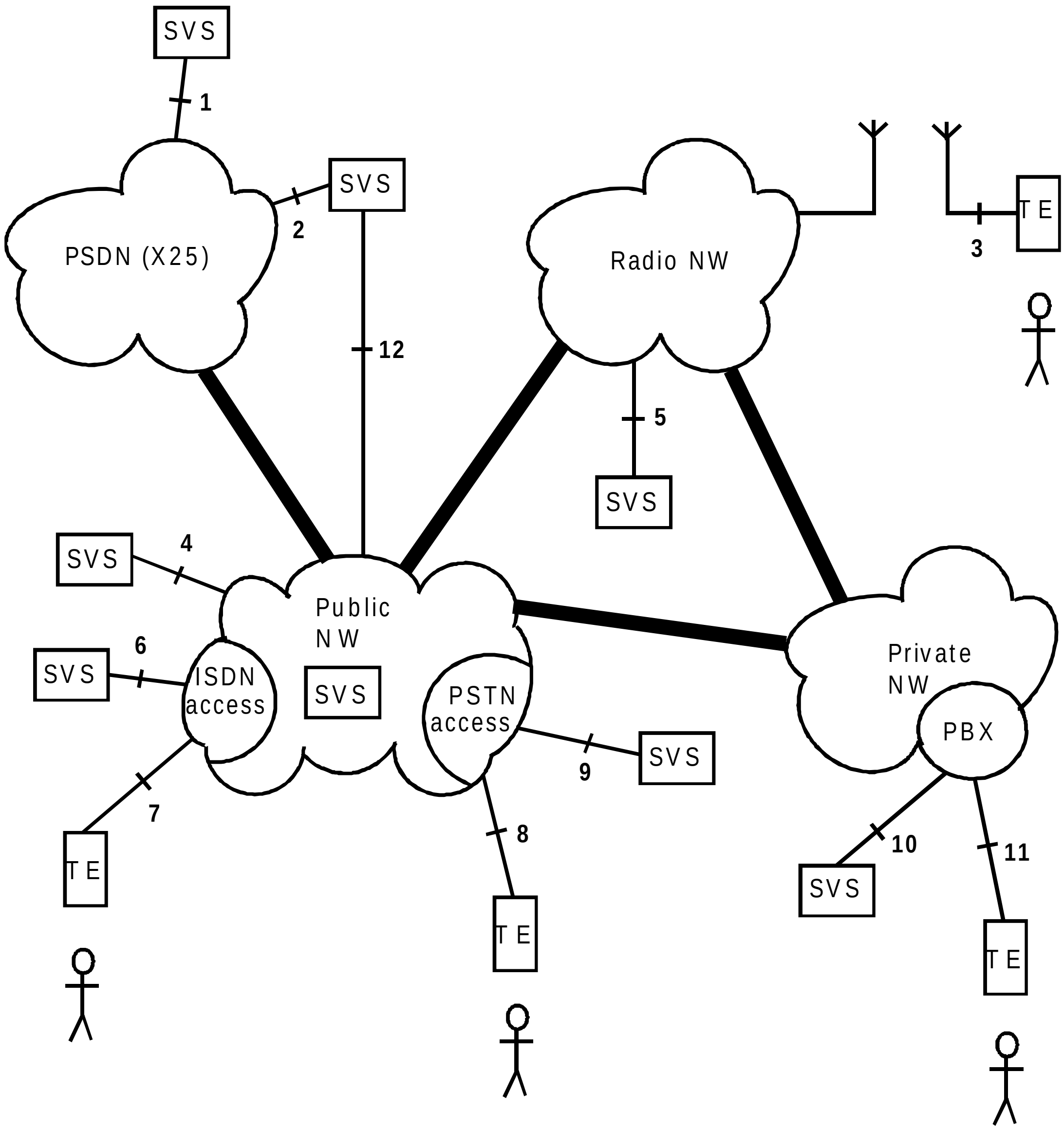
For the data interchange, high speed transfer is expected and the interchanges needs to be conveyed through heterogeneous networks linked by gateways and through various kinds of access by various different type of SVS platforms, therefore, functional standards (or profiles) need to be used to allow end-to-end interworking between any SVS platforms. A typical example can be seen in figure 2 for the connection <10-2> passing through a PBX in a private network, then through the PTN network and then through a Packet Switched Data Network (PSDN) (ITU-T Recommendation X.25 [23]) network.

There is also a need to interchange data through the Public Telephone Network (PTN), and through analogue PSTN access at the highest possible speed by using a modem allowing fast transfer (e.g. V.32bis/V.42bis [24], [22] modem). This is mainly because the PTN is the more widespread network (especially in the countries where ISDN, ITU-T Recommendation X.25 [23], networks are not available yet) and the use of a modem is a cost effective solution which is even available on Personal Computers (PCs).

Another quite common interconnection mode is the interchange through the PTN and via the PSDN as e.g. ITU-T Recommendation X.25 [23] to access to a SVS connected to the PSDN. (see the connection case <8-2> in figure 2).

The interworking with radio cellular networks such as for example the GSM and the other main networks is of great importance since voice messaging is an important feature to mitigate the effects of user mobility and the very frequent impossibility of reaching a user who is out of the covered area or simply off line.

In the countries where the ISDN is significantly available a connection from an ISDN SVS to an ITU-T Recommendation X.25 [23] SVS can be also contemplated through the ISDN B- or D-channels.



Description key: TE = Terminal Equipment
SVS = Stored Voice Server
<x-y> means communication between the interfaces x and y
<x-N> means communication between the interface x and a server which is part of the Network

<1 -2 > :	Data interchange between two stored voice servers (SVS) through a public packet switched data network.
<3 -5 >:	Access to a SVS using a mobile telephone through a radio telephone network.
<4 -8 >:	Automatic outgoing call from a SVS to a user's terminal equipment (TE).
<10-1>:	PSTN user TE access to a SVS connected to PSDN (e.g. X.25 [25]) through PSTN/PSDN gate-way.
<6 -11>:	ISDN stored (SVS) to a PBX user's TE.
<7 -N >:	ISDN user's TE access to a SVS being part of the network.

Figure 2: Schematic diagram of possible interconnections

Eventually, for some specific needs an access through various LAN or MAN can be also foreseen may be only through some gateway as LAN/ISDN or LAN/X.25.

All combinations cannot be offered for optimisation reasons but a sufficient set of profiles should be available to cover the main needs to interconnect through the main available network in a country.

7.5.5 Intelligent Network (IN) architecture

A SVS can be integrated in the network and directly connected to the signalling system of the network with SS7 [10], [12], [13] for the public network and with DPNSS, for some private networks or with some proprietary internal buses. The implementation of services in the network is growing rapidly and a need to separate the switching functions of the network and the services functions is becoming dramatically necessary. It is the reason why the IN architecture is now emerging rapidly. Since the IN architecture has to be implemented in the existing networks, the services will be in a first stage supported in the form of peripherals.

The IN concept is described in detail in the new ITU-T (former CCITT) Q.1200 series of Recommendations on Intelligent Networks [25].

IN architecture is applicable to a wide variety of networks including but not limited to, Public Switched Telephone Network (PSTN), Mobile Cellular Network (MCN). Packet Switched Public Data Network (PSPDN), and Integrated Services Digital Network (ISDN) - both Narrow band-ISDN (N-ISDN), and Broadband ISDN (B-ISDN). IN supports a wide variety of services, including supplementary services and utilises existing and future bearer services (e.g. those defined in N-ISDN and B-ISDN contexts).

IN is an architectural concept for the operation and the provision of new services which is characterized by the extensive use of information processing techniques, the efficient use of network resources, modularisation and reusability of network functions, the creation of integrated service and their implementation by means of the modularised reusable network functions, flexible allocation of network functions to physical entities, portability of network functions among physical entities, standardized communication between network functions via service independent interfaces, service subscriber control of some subscriber-specific service attributes, service user control of some user-specific service attributes and standardized management of service logic.

In this context a service subscriber is a person or entity who, or which, obtains a service from a service provider and is responsible for the payment of the charges due to that service provider. A service user is a person who has access to, and makes use of, services.

A key objective of the IN is to provide service independent functions that can be used as "building blocks" to construct a variety of services. This allows easy specification and design of new services.

A second key objective is network implementation independent provision of services. This objective aims to isolate the service from the way the service-independent functions are actually implemented in various physical networks, thus providing services that are independent of underlying physical network infrastructures. The main objectives of the network implementation independence are, services that use distributed network functions in various ways, services that can span several networks and are independent of specific implementation in the network, services that are independent of technological developments and evolution in the network infrastructure, so that physical networks can evolve without affecting existing services, and physical elements in such a network can be procured from different vendors.

This concept which seems very attractive for the services providers wishing to support their service on a very powerful network infrastructure, without taking the risk of investment in a stand-alone platform which could become rapidly obsolete. But it needs to be noted that the IN architecture will be implemented step by step in various phases taking into account the existing infrastructures of the various network and, therefore, will not reach the final objective for several years.

Nevertheless, in the framework of the IN architecture, the "IN service plane", will probably become, in the near future, a captive platform for the provider of services to offer Multi-Application Voice Processing (MVP) services, by using in an intermediate stage some intelligent voice messaging platform more or less closely interconnected with the network as a peripheral.

7.6 Terminal equipment accessible by voice services

Voice services can thus output information to a whole range of terminals, not only to those with voice telephone facilities, such as:

- ordinary telephones;
- facsimile machines;
- pagers;
- Videotex servers;
- Videotex Terminals;
- telegraph terminal equipment;
- ISDN terminals;
- personal, mini or mainframe computers;
- other MVP servers;
- peripheral equipment in the Intelligent Network.

8 Interchange protocols

8.1 General

The progress of technology, which has brought more powerful mini- and micro-computers with larger memory capacity for lower prices, and the implementation of advanced voice coding/decoding/recognition algorithms, has led to more and more stored voice systems being set up.

For several reasons such as efficiency, price and organisation, it is often necessary to interconnect these systems. For instance, a multinational company operating a voice mail system in a number of countries may want a global system, allowing exchanges of voice mails throughout the whole company, in a transparent way. Alternatively, a service provider setting up several voice information servers may want to be able to distribute information updates between the servers in an automatic way.

Furthermore, it should be possible, and simple, to interconnect equipment from various manufacturers, independently of network connection or the format of any voice information.

More precisely, the following list presents several identified requirements for a stored voice system interconnection solution:

a) interoperability;
   - independent of equipment to be interconnected;
   - independent of information to be exchanged;
   - independent of lower network layers;

b) evolvability;
   - allow for future/unforeseen extensions (e.g. multimedia);

c) security;
   - information origin authentication;
   - secure access;
   - content integrity;
   - content confidentiality;
   - non-repudiation;

d) general capabilities;
   - directory services;
   - easy management;
   - easy testing;
   - support for billing.

These requirements have already been addressed, partly or fully, in various documents, some dealing specifically with store voice systems others being more general. It is possible to classify them into two categories, analogue and digital.

The main differences between usual voice communications and stored voice interchanges are the additional needed control information (mailbox addresses, security attributes, etc...) and the fact that transmission is not done in real-time. Transmitting voice in an analogue form has been the simplest and most cost-effective way for a long time. The main drawback, when considering the interchange of stored voice messages, is the poor capability for carrying signalling and controlling information.

On the other hand, digital transmissions allow for a rich set of capabilities. Furthermore, because transmission is not done in real-time, the transmission of digitized voice data can appear as a special case of file transfer.

The following presents a brief description of several activities, complete or ongoing, in the fields of both analogue and digital interchange protocols. Noting the above discussion, a short analysis is then presented.

8.2 Analogue interchange protocols

8.2.1 AMIS-Analog

The Audio Messaging Interchange Specification (AMIS) process was initiated in mid-1987 by a US company in conjunction with two voice mail system vendors. Their initial concern was to be able to connect dissimilar systems within a company network. It was soon realised that this effort could interest other manufacturers, system and service providers and users, which were then contacted and a co-operation forum was set up. The resulting number of participating companies was 36 at the beginning of 1992.

The AMIS-Analog protocol [13] defined by this process provides a mechanism for transferring voice messages between two stored-voice systems in a fully analogue manner. Signalling is done using DTMF and the actual message is transmitted in analogue form. Message formats are defined for originator and recipient identification, mailbox addressing and reply messages.

The AMIS-Analog protocol [13] is designed for use on the PSTN. It does not provide any complex capabilities such as delivery notification, time-stamping, security, etc...

The AMIS-Analog protocol [13] description includes DTMF requirements, test facility requirements, call set-up, description of message formats, message exchange description and interworking with the AMIS-Digital protocol [14], its digital counterpart.

8.3 Digital interchange protocols and related protocols

8.3.1 CCITT Recommendation X.400

The CCITT X.400-X.420 series of recommendations [16] define the overall system and services of a Message Handling System (MHS). Exchanged data may be text files, facsimile data, voice data or other data.

More precisely, CCITT Recommendation X.400 [16] describes the system model and elements of service of the message handling system and services. The MHS can be constructed using any network within the scope of Open Systems Interconnection (OSI), using presentation layer services and services offered by other, more general, application service elements. The message transfer service is application independent.

The MHS functional model includes following elements:

- the user sending a message (named the originator) prepares it with the assistance of a User Agent (UA). The UA is a process which interacts with the Message Transfer System (MTS) or a Message Store (MS);

- the MTS delivers the messages submitted to it, to one or more recipient UAs, MSs, or Access Units (AUs) (gateways to other telematic services). The MTS comprises a number of Message Transfer Agents (MTA);

operating together, in a store-and-forward manner, the MTAs transfer messages and deliver them to the intended recipients. The primary purpose of the MS is to store and to permit retrieval of delivered messages.

The CCITT Recommendations [16] define the MHS overall architecture (which includes the security model), conformance testing, encoded information type conversion rules, access and transfer protocols and message structure.

Some information conversions such as ASCII to group 3 facsimile exist. Voice to other format and other format to voice conversions are listed as "impractical". Voice-to-voice conversion is listed as "for further study".

8.3.2 CCITT Recommendation F.400

The CCITT F.400 series of Recommendations [26] describe public message handling services. CCITT Recommendation F.400 [26] is identical to CCITT Recommendation X.400 [16]. Other Recommendations describe naming and addressing for public message handling services, the general, operational and quality of service aspects of the public international message transfer service, the intercommunication with public physical delivery services, and the intercommunication with the telex and the teletex services.

8.3.3 CCITT Recommendation X.440

CCITT Recommendation X.440 [15] defines:

the voice messaging system, a form of message handling tailored for exchange of voice encoded information;

two kinds of information objects: voice messages and voice notifications. Voices notifications are used to report voice message delivery status;

CCITT Recommendation G.721 [20] (32 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)) voice encoding algorithm as the default algorithm. Other encoding algorithms may be specified in the corresponding header field attached to a voice mail.

Various other header fields are provided: originator and recipient, along with expiry time, importance, sensitivity, language, etc...

Objects of type different from voice information can be attached to a voice message.

Also defined are the operations that can be performed in the voice messaging environment: originate voice messages, originate voice notifications, originate probes, receive voice messages, receive voice notifications, receive reports and change management parameters (e.g. auto-discard, auto-acknowledgement, auto-forwarding).

8.3.4 CCITT Recommendation F.440

The CCITT Recommendation F.440 [27] specifies the general, operational and quality of service aspects of the public international Voice Messaging service. Elements of service specific to Voice Messaging are listed: attendant-assisted delivery, forwarding voice message indicator, replying voice message indicator, etc...

8.3.5 CCITT Recommendation X.500

The CCITT X.500-X.521 series of recommendations [28] describe the directory capabilities required by OSI applications, OSI management processes, other OSI layer entities, and telecommunication services. Among these capabilities are those of "user-friendly naming" (objects can be referred to by names easily remembered by human beings), and "name-to-address mapping" (dynamic bindings between object names and their location).

These CCITT Recommendations [28] define the Directory Information Base, composed of directory entries, each of which contain a collection of information (attributes) on one object. Also defined are the Directory service, Directory distributed operations, Directory protocols and authentication services.

8.3.6 ITU-T Recommendation F.500

CCITT Recommendation F.500 [26] describes the requirements for public directory services, and the associated service features. It specifies naming aspects, describes operational issues and quality of service aspects.

8.3.7 ISO 7498, Part 2

ISO 7498, Part 2 [29] provides a general description of the security services and related mechanisms, which may be provided by the OSI Reference Model. It also defines the positions within the OSI Reference Model where these services and mechanisms may be provided.

The following services are presented: authentication, access control, data integrity and non-repudiation.

The following mechanisms may provide these services: encipherment, digital signature mechanisms, access control information bases, authentication information, capabilities, security labels, sequence numbering, time stamping, traffic padding, security audit trail, etc...

8.3.8 AMIS-Digital

The AMIS-Digital protocol [14] is the counterpart of the AMIS-Analog protocol [13]. It provides a mechanism for transferring voice messages between two stored-voice systems. The protocol relies on the CCITT X.400 series recommendations (1984 Red Book version) for the Application layer. It specifies which CCITT Recommendations are to be used for each other OSI layer (X.409, X.225, X.224, X.25 Packet, X.25 [23] LAPB, V.32 or V.35).

The document also describes how AMIS-Digital features are mapped to the CCITT X.400 series of recommendations [16] (sending, receiving, delivery notification handling, error handling, forwarding, transfer). A possible testing procedure is presented. Finally, interworking with AMIS-Digital is described.

Specific information is given related to speech encoding algorithms, speech quality trade-off, algorithm evolution and format conversions.

8.3.9 Asynchronous Protocol Specification (APS)

A group of vendors led by Norwegian Telecom and a US company joined in an alliance in 1992, to develop and promote a standard means for the CCITT X.400 series of recommendations [16] communications using standard dial-up telephone networks. The Asynchronous Protocol Specification (APS) proposed standard ("Asynchronous Protocol Specification-APS") [30] is intended to support all OSI applications, giving them a way to exchange data rapidly and reliably using the existing switched telephone network. It is specifically targeted for use with the CCITT X.400 [16] series of recommendations on messaging and the CCITT series of recommendations X.500 [28] directory services. The APS protocol [30] relates to OSI layers 1 to 3. APS future versions may apply to mobile radio and satellite-based communications.

8.4 Questionnaire replies - use of interchange standards

Answers to the questionnaire have already been analysed (see Clause 5). Table 1 below presents figures specific to the interchange protocols.

Table 1: Questionnaire replies

Protocol	Used	Not used
AMIS-Analog	8	19
AMIS-Digital	6	21
Other	6	21

The interchange question was not applicable to four of the organisations which replied (e.g. consulting firms, etc.).

AMIS-Analog was mentioned 8 times in the answers to the questionnaire, compared to 6 times for AMIS-Digital. The analogue way to interchange voice mails seems to be the most used, perhaps due to its simplicity. However, caution needs to be exercised with this conclusion due to the low number of replies to the questionnaire.

File Transfer Access and Management (FTAM) and CCITT Recommendation X.440 [15] were listed in the "other" category. Use of FTAM is a good example of the fact that voice data can be interchanged as any other data. Reference to CCITT Recommendation X.440 [15] seems to indicate that companies are already looking at standardization work being performed by official bodies.

8.5 Discussion

8.5.1 Analogue protocols

The AMIS-Analog protocol [13] seems to be the only analogue interchange protocol to be available to date. Used by several system providers, it allows for cost-effective and simple solutions. Its main limitations are the low number of control messages that can be transmitted along with stored voice data, preventing sophisticated applications from being set up (delivery notification, time-stamping, message enciphering, multimedia etc.).

This technology can be seen as obsolescent and no further work appears to be needed in this field.

8.5.2 Digital protocols

As described at the beginning of this Clause, digitized voice data can be treated as being no different to "normal" data that can be interchanged using usual file transfer protocols. Of course, semantics specific to stored voice data need to be attached to interchanged data. This is usually addressed only at OSI application layer level.

In the digital domain, there was no recommendation or standard endorsed by an official body, related to voice messaging, previous to CCITT Recommendation X.440 [15], which describes the voice message system, at the application level. The AMIS-Digital protocol [14] is more profile-oriented, describing the various protocols to be used for each OSI layer. It needs to be noticed that the AMIS-Digital application layer does not follow CCITT Recommendation X.440 [15], as work on it started before the CCITT Recommendation. AMIS-Digital also restricts itself to the use of ITU-T Recommendation X.25 [23] PSDN, and other given standards for each one of the other OSI layers.

8.5.3 Asynchronous links

Considering the fact that the CCITT X.400 [16] series of recommendations applications growth was restricted due to the lack of support for asynchronous connections, the APS Alliance decided to create and promote APS. One of the APS aims is to allow for delivery of easy-to-use products, supporting new technologies (V.42 error-correcting modems), point-to-point connections, Packet Assembly/Disassembly (PAD) accesses, etc... APS seems a good candidate to possibly enlarge the scope of the AMIS-Digital profile [14], for instance.

Use of asynchronous connections may possibly increase, thanks to efforts like the APS one. It should also be noted that modern modems are now able to exchange data at higher rates: 14 400 bit/s for CCITT Recommendation V.32bis [24] modems, possibly multiplied by a factor of 4 with CCITT Recommendation V.42bis [22] compression. V.Fast will further increase the transmission rates. This could lead to more and more applications using the widely available PSTN for voice data interchange. Not a lot of work has been done in this field to date, apart from APS, and a few file transfer protocols (Kermit, XModem, YModem, ZModem, etc.) in use for a long time, mainly among personal computer user community. However, these later ones were not developed bearing in mind any idea of compliance to international standards.

The lower OSI layers do not need to know they deal with voice data. Only layer 7, the application layer, is in charge of handling the specific meaning of voice data. Being able to interchange voice data over various networks, between equipment of various manufacturers, means first that lower layers are able to dialogue in the right way. This is usually granted by the use of functional standards or profiles. These functional standards are already defined for the existing standards.

8.5.4 Non-OSI models

OSI is not the only model in use in the network world. Transmission Control Protocol/Internet Protocol (TCP/IP), developed by the US Department of Defence, has a wide acceptance in the US, and provides some competition to OSI in Europe. Other protocols exist, which are proprietary, like SNA from IBM, DECnet from Digital (the latest version being OSI-compliant), etc... Nothing prevents stored voice interchange applications from using these protocols, as long as the right interfaces and gateways exist, when the use of different kinds of networks is considered.

8.5.5 Intelligent Networks

Intelligent Networks (IN) availability could change the way service providers design and implement their services. In the case of stored voice data interchange, the IN could offer the interchange functionality as a network service. This would prevent the stored voice service provider from having to implement the interchange part. However, use of the IN interchange functionality would still require a defined protocol.

8.5.6 Multimedia

The ability to handle and exchange documents containing various kinds of information (voice, text, graphics, video, etc.) is expected to become more and more important. One possible consequence could be that stored voice applications would need to interchange not only voice data, but also more complex data. The CCITT series of recommendations X.400 to X.420 [16] can handle various data formats. Although there is ITU-T work in the H.200 series of Recommendations it should be noted that the multimedia field is currently far from being fully standardized. Data and application portability, data replay/record/conversion, synchronisation between video replay and other data types, etc., are still under investigations. This work is not specific to stored voice services.

Within ETSI, standards exist for inband signalling procedures for audiovisual terminals (ETS 300 143 [31]) and for the frame structure and associated syntax for inband signalling (ETS 300 144 [32]).

In the USA, a new standards writing group entitled the Universal Messaging Interoperability Group (UMIG) has recently been set up to foster the development of universal messaging to fulfil the needs of users "who will soon be demanding true multimedia messaging capabilities". This demonstrates how such fora are able to work rapidly in advance of the market.

8.5.7 Security

Security of stored voice data interchanges can be performed using the existing techniques: controlled access to computer files, user validation to prevent unauthorised people from logging in on one of the stored voice servers. The interchange part of the service also needs to be protected. For this purpose, the CCITT X.400 to X.420 [16] series of recommendations includes a number of protocols related to security: proof of delivery, content integrity check.

8.5.8 Discussion

It is clear that there is already much activity on interchange protocol standards and that no new work in this field needs to be started within ETSI. Apart from ETSI work on multimedia terminals (ETS 300 143 [31] and ETS 300 144 [32]) and some European input to work on the APS protocol [30], most of the voluntary standards writing fora are American led. It would be to their advantage if more European companies took part in these standards writing activities.

In order to encourage companies in Europe to take part in non-European voluntary standards writing fora it is recommended that:

Recommendation 2

Support should be given to participation in non-European voluntary standards fora that are relevant to stored voice services.

9 Performance of stored voice services

9.1 General discussion

This Clause of the ETR discusses the general question of the performance of stored voice systems in human factors terms, including user interface design and the present state of relevant standardization. The findings from the part of the questionnaire dealing with the user interface for existing systems (see table 2) have been taken into account before drawing any general conclusions about standardization needs in this area.

Most, if not all, stored voice systems have been designed around a convenient existing platform - a public network. The functionality provided by the system is available to an end-user via a normal telephone. This arrangement has advantages and limitations, both in terms of performance and usability.

The technical performance of stored voice services may be assured by adherence to whatever technical standards, regulations or guidelines are applicable to the specification (see Clauses 7 and 8) and by appropriate normal quality assurance procedures for software. End-to-end performance technically may, in addition, have to take into account the performance of any peripherals or end-user terminals. However, the technical performance of such a system is not the only criterion and overall performance should also take into account its usability.

Usability as a concept is made up from a number of components. Very few technical standardization bodies recognise the need for some sort of standard which defines usability, but it will be defined in ISO 9241, Part 1 [34], which deals with guidance on measuring usability for software systems, as follows:

> "**The effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments"**

Effectiveness - is the accuracy and completeness with which specified users achieve specified goals in particular environments.

Effectiveness of a stored voice service is a measure of how well the system meets its stated objective, from a users point of view. This is partly a function of how the service is described, but also how the service is perceived by the user. A user can use the wrong service, or the right service badly, and get for him what is an ineffective result. The end result may be just as likely to be a marketing problem as a user interface problem.

Efficiency - considers the resources expended in relation to the accuracy and completeness of the goals achieved.

The efficiency of a stored voice service is related to the speed with which transactions may be completed, and the number of errors that a user makes before completion (or giving up). Clearly speed and accuracy can to some extent be traded, and are also directly related to users' experience. In addition to learning and experience, in this case the physical user interface can also significantly effect usability. Some keypads (for input) are more difficult to use than others (too small, inadequate spacing, poor or illegible graphics, etc.) and the quality of the voice output, especially prompts, can influence the speed and accuracy with which a user can respond. Speech quality is a function not only of the telephone network, the end-user terminal and the environment (noise, etc.), over which the system developer has no control, but also the type of speech recording, coding and processing, over which he has some control.

Satisfaction - is the degree of physical comfort and acceptability of a system to its users and other people affected by its use.

Satisfaction expresses the well-being felt by a user who has used a system that has performed as expected, and who has completed a task quickly and accurately. Positive satisfaction is a good indication of the usability of the system, as it implies that most of the elements have been designed to promote user friendly dialogue. Negative feelings about a system may be invoked by any of the components which lead to user mistakes. Many users react negatively the first time they encounter stored voice systems just because they are unable to speak directly to another person.

Any measurement of usability would attempt to quantify the components of effectiveness, efficiency and satisfaction by appropriate performance measures, that have always to be qualified with reference to the users of that system, and measured under specified conditions. ETR 095 [33] provides a guide to usability evaluation, which would form the basis for quantifying the usability of stored voice services and the availability of such a measurement would assist purchasing decisions in an open and competitive market.

In order to assist purchasing decisions in an open and competitive market it is recommended that:

Recommendation 3

Methods should be developed to quantify the usability of stored voice services.

The users of stored voice systems may be a defined target group, if the system is specialised, or more usually will be the general public, i.e. undefined. The experience of individuals within this broad category will have an important effect on usability, related to learning and the learnability of the system. When a stored voice system is open to access from anyone, the chances are that no prior learning will have taken place or be possible, although there are some exceptions to this. Where the target group is an ordinary telephone subscriber, then the procedures may be explained in the pages of a telephone directory. The customers of a bank may receive a general brochure about telephone banking procedures. In either case, the procedures need to be clearly spelled out, as well as the rules applying to the dialogue, so that a user can expect a "machine service", and assimilate the general principles and learn quickly from specific examples. However, provision of such material does not guarantee a user will read or learn it, so that the use of stored voice systems should be as intuitive as possible. Some systems do provide adequate explanation and help on-line, but there are obvious limitations to its extent.

Users may be first-time, one-time, or experienced. Probably the majority will fall into the first two groups, as even regular users may not use the system at regular enough intervals to become experienced. Such occasional users will probably always require a full introductory statement and full explanations and help available. Experienced users, on the other hand, should be able to use short cut methods, such as not having to listen to a complete menu before making a selection from the available options.

If usability is properly considered, the availability of on-line explanation, help systems, and especially a "repeat" command, will materially assist all categories of user and should, therefore, always be provided.

9.2 The user interface

As mentioned above, the user interface has strengths and weaknesses. Its strength lies in its universality, its simplicity and its robustness.

Stored voice systems mostly make use of an interactive dialogue with the user, where a command is entered via control keys or voice. Feedback is provided by the system together with further instructions or prompts for the next command, and so on. Efficient systems deal with error correction in the same way. Choices available to the user are often presented in audio menus, with the user entering a number appropriate to his choice. Clearly from the usability point of view, the way that instructions, prompts and feedback generally (especially error correction strategies) is critical and requires careful design to allow for the wide range of user abilities.

Telephones are almost universally available. Most services make use of DTMF signalling via the 12 key pad, but some network-based services can also accept pulse signalling from older phones. In future, more elaborate phone-based interfaces may make use of letters on the numeric keypad as these are being re-introduced by a number of European Union countries and are now widely featured on GSM and other mobile phones.

Increasingly, the user dialogue may contain voice recognition elements, although presently having only limited recognition capability. Currently, the only way of having a comprehensive dialogue with stored voice services from a non-DTMF phone is through automatic speech recognition, with a minimum vocabulary. Improved, more reliable, voice recognition algorithms are developing rapidly, which will lead to an increase in their application generally for stored voice services in the future, and service applications having a much bigger vocabulary are likely to become more common. Usability of such systems is likely to be improved, compared with their purely phone-based minimum Man Machine Interface (MMI) counterpart.

Almost everyone is familiar with the phone and how to enter digits with the keypad. Payment for services, whether they are free, charged at normal rates or at "kiosk" or premium rates, presents no problem for most users. Even when services are available from public payphones, if they are not free, they can be paid for in the normal way with coins, phonecard or credit card. Stored voice services may also be faster than an operator based counterpart and automatic services may give a quicker reply to the user than operator based systems. On these criteria, stored voice services may be judged to be both efficient and effective.

However, a weakness in such services also stems from the simplicity of the interface for the user. Four main aspects can lead to both inefficiency and poor satisfaction with such services, including:

a) the telephone provides audio output only;

b) information to the user is provided serially;

c) the telephone has only 12 keys;

d) users generally have no supporting material.

The audio output has to provide not only the information or service the user actually wants, but all the feedback, prompts and error correction messages necessary for successful completion of the user dialogue. The user has no control (usually) over the volume or quality of speech or tones from the telephone, nor the noise environment in which he uses the service. Users with hearing problems (not necessarily just loss of hearing) and also those with speech impediments may find particular difficulty using stored voice services.

The serial nature of telephone audio output in particular puts a very heavy demand on human short term memory. Users will normally rely on the service to provide on-line instruction and help for the dialogue, but human memory limitations restrict the amount of help or prompt material ("chunks" of information), that can be offered at any one time, which in turn slows down the dialogue. Consequently, this places a heavy demand on design of the service but especially design of the dialogue.

Nearly all the information a user provides to the system is numeric, or numerical coded alphabetic, entered through 12 keys. At present, outside the USA, few fixed base phones have any alphabet letters assigned to number keys, although draft revision of CCITT Recommendation E.161 [35] exists and it is being re-introduced by a number of countries. ISDN phones may come equipped with a full alphabet keyboard as well as some feature phones or special terminals (e.g. MINITEL) However, radio mobile phones (GSM, etc.) are being used in increasing numbers, with alphabet letters assigned to the keypad, so that service providers may design future services based on this type of input. An alphabetic input of this kind also requires an additional command to indicate that an alpha input is to follow, or which number key is being pressed and the letter position on that key. A number of implementation methods is possible, including the current GSM method, where double or triple presses on a key give the appropriate alpha - the accuracy of which is critically dependent on timing by the user. Another way is to specify each letter by two digits, the first being the number key and the second the position of the letter on that key. Such a method needs, in addition, a shift key to signal the change of function, or it could be a "soft" change, i.e. context dependent. No attempt appears to have been made to standardize any of the possible methods.

Moreover, many fixed base phones and all mobile phones have the keypad built into the handset. This means that the user needs to remove the handset from the ear and possibly reverse it many times during the dialogue. This leads firstly to extended response times, and secondly to probable increased error in keying. Voice responses also may suffer degradation especially in terms of speed of input or response to prompts. These place users of such interfaces at a severe disadvantage compared with "ordinary" phone users, and service designers should take account of this when designing dialogues, especially with respect to length of any time-outs and error correction routines in help systems.

In order to assist the growth of a free market in stored voice services it is essential to improve the consistency in their operating procedures it is recommended that:

Recommendation 4

Guidelines should be developed for the user procedures and dialogue in stored voice services, especially that for entering alphabetic characters on telephone sets with alphabet letters assigned to numbers.

The use of displays, now common to all mobile phones, is a possible way service designers may improve the design of user interfaces and dialogue especially for those users with speech and hearing difficulties, but is outside the scope of this ETR.

Nevertheless, designers should be aware that ETSI standards are proposed for display and messaging in the field of Calling Line Identity (CLI) services, and that others (e.g. Bellcore Analogue Display Services Interface) exist, and they should ensure that, as far as possible, there is a relationship between spoken and written messages when they are options on the same system.

Support material available to users of stored voice systems is likely to be minimal. It seems to be widely assumed by service providers (not necessarily the developers) that users will be familiar with data entry procedures, or that they are such experienced telephone users that entry routines based on the 12 key pad will be easily learned and executed. In fact, because most voice telephony does not contain interruptions, breaks or prompts, the majority of first time users will not find an interactive dialogue easy (a similar problem may be encountered when attempting to use network supplementary services). Specialist service providers, such as banks, may send comprehensive explanations and instructions to their customers, sometimes a summary procedure card may also be provided. In such cases the user will usually be reasonably motivated and a satisfactory outcome may result. A user encountering a voice messaging system for the first time, on the other hand, without pre-knowledge, may react in the same way as when hearing a telephone answering machine - terminate the call.

The consequence is that usability of a service depends to a large extent on prior knowledge of some sort, as well as an explicit and comprehensive voice welcome and step-by-step instructions and prompts on-line. Standards relating to such issues are not considered part of ETSI's terms of reference, neither are these issues addressed in the existing guideline documents but could, if available, materially assist service providers to improve both the uptake and use of their systems.

To improve the uptake in the use of stored voice services it is recommended that:

Recommendation 5

Guidelines should be developed for the preparation of introductory and user procedural text and voice messages for use in stored voice services.

9.3 Voice aspects

Stored voice systems, by definition, use speech as a primary means of input and output for messages, information and instruction. Voice from the user may also be used for command input, and was identified in the questionnaire. Voice activated systems, especially if demanding only voice input, may be difficult for first time users, and will also be denied to persons with speech impairments of some kinds. Speech as output is universally used for introductory or greeting messages, instructions, prompts and help, and for audio menus.

The choice of speaker and style of speech for recording purposes, together with the quality of recording and processing can have a marked effect on the usability of the output, because of its influence on intelligibility for the user. Persons with hearing impairments will be especially effected. Aspects of speech encoding and recombination (concatenation) for both output and text-to-speech conversion are well discussed in some of the guideline documents (see ISO/IEC JTC1/9018 N.4227 [8] and the France Telecom "Recommendations aux partenaires des services vocaux Téléphoniques" [36]) and transmission and sound reproduction aspects (frequency bandwidth, distortion) are the subject of telecommunications standards or recommendations. Quality of speech, including recommended limits for distortion, is discussed in such documents, but again in the context of telecommunications generally although there is a new ITU-T Recommendation P.85 [37] dealing with subjective test methods for assessing the quality of speech voice output devices such as voice servers.

Speech, as command or control input, involving Automatic Speech Recognition (ASR), is not yet commonly introduced in stored voice systems, and is generally restricted to numbers and simple commands, such as STOP, START, etc., and the numbers. One advantage lies in its being able to supplement DTMF signalling, or substitute for it completely in areas where it is not available. Another is its potential flexibility. ASR in stored voice systems has great potential once the general problems of speech recognition, especially for connected or continuous speech, have been solved and truly speaker-independent recognition at very high accuracy becomes possible. Application development in this area should therefore be encouraged. Until that time there may be an advantage to promote standards for vocabulary, at least. The use of a speaker-dependent system may have some benefit in some security sensitive applications, allowing a voice to be recognised and obviating or supplementing the need for a user to remember his ID or PIN number.

To assist manufacturers to make more effective use of automatic voice recognition it is recommended that:

Recommendation 6

Support should be given for the preparation of guidance on automatic speech recognition for stored voice services.

A significant problem for ASR systems is that of interference from other speakers or from general ambient noise, which already interferes with voice output. This problem may be solved as telephone network operators improve their systems generally, by using noise rejection circuits or multi-microphone input. It is not a problem special to stored voice systems, but could delay their general introduction and use. The common use of ASR in the future will bring with it the important requirement for a user to be able to choose his language. In this event, it will clearly be desirable for a common European, if not world-wide, standard to be set for telephone numbers for this purpose. Country codes have been proposed for this purpose, but are not suitable for multi-lingual countries. A standard already exists in the ISO Code for the representation of names of languages. ISO 639 [38], is a possible starting point for development as a necessary ancillary to more widespread use of ASR systems in stored voice services.

Therefore to encourage the provision of multi-lingual stored voice services it is recommended that:

Recommendation 7

A standard for language identification codes should be developed.

9.4 User standards

From the sample of questionnaire replies analysed, no standards have been referred to which define any requirements for the overall performance or usability of stored voice services. Replies to the questionnaire referred directly or by inference to only four standards or guidelines which dealt with any part of the user interface.

The first is the almost universally accepted CCITT Recommendation E.161 [35] for layout of the number keypad, and for the assignment of letters to numbers on a keypad one version of which is likely to be adopted by ISO as a standard.

For other parts of the interface or user procedure, three guideline documents exist, although not yet in a published standards form. The first of these, now in the form of a proposed Draft International Standard, has been produced by a joint technical committee of ISO/IEC as ISO/IEC JTC 1/SC 18 N 4227 [8], User Interface to Telephone-based Services: Voice Messaging Applications". This document is intended to provide users of voice messaging systems with a consistent mode of interaction in a way that is independent of the underlying system implementation. It provides guidelines for voice or DTMF input, system output, time-outs, system response times and basic use of the ★ and # keys. It is written for application to Call Answering, Voice Mail, Voice Bulletin Boards and Message Delivery systems. Since it is generic, and intended to cover such a wide variety of applications, it does not deal with all of the specifics for stored voice services, except in an Appendix when replay, skip forwards and backwards are specified on number keys for voice recordings. It contains no discussion of end-to-end quality or usability aspects, nor does it propose any conformance or compliance testing.

Table 2: Responses to questionnaire - Question 6

QUESTION 6 - User Interface							
Systems described	Procd. standard	Controls			Prompts		
		keys	* - #	voice	tones	voice	text disp
TRANSVOX	CNET	✓		✓	✓	✓	✓
PTT SERVICES	PTT	✓		✓	✓		
VOICE MAIL	?	✓		✓+ hook		✓	✓
PHONE MAIL (USA)	VMUIF			flash			
MANAGED ANSWERING SVCS	BT Style	✓	✓		✓	✓	
INNOVOICE	?	✓	✓		✓	✓	
Phone Mail	?	✓	✓			✓	
GENLOGUE, DATALOGUE	CCITT	✓	✓	✓	✓	✓	
PDL _Progr.Defn.Lang.	?	✓	✓	✓	✓	✓	✓
DIVAPHONE	CNET	✓	✓	✓	✓	✓	✓
VMS	?	✓	✓	✓	✓	✓	✓
VM SERVICE	Co.Std (VMUIF?)		✓			✓	

The Information Industry Association in the USA sponsored the Voice Messaging User Interface Forum (VMUIF), which has issued a "Specification document" [39] dealing with a number of interface aspects. The Forum's objectives are to provide users of voice messaging systems with a consistent mode of interaction, to define common elements of the services from the users point of view, to facilitate the offering of voice messaging services in a way that is independent of the underlying system and to promote future consistency and user transparency as the user interface becomes defined. Recommendations in the specification document (which may have de-facto standard status in the U.S), fall into three areas:

1) general guidelines and principles that are recommended for good design of user interfaces for voice messaging;

2) a set of core features that all voice messaging systems should include; and

3) keypress assignments for voice messaging systems and functional block diagrams for various menus.

Only one reply to the questionnaire alluded to this specification, and one other may be using it.

The replies to the questionnaire included one other document which may be considered as more application specific to the design of the user interface for stored voice services. France Telecom has published a set of User-friendly recommendations for Voice Services Designers [36]. Much of this is devoted to voice quality, where it deals with many of the issues raised above, including references to an existing method for the subjective determination of a standard for transmission quality (ITU-T Recommendation P.80 [40]), and others from the field of telecommunications generally. It also includes harmonization issues, such as key assignments, menu design, time-outs, etc., for voice message applications.

In addition to these guideline documents, ETSI has produced ETR 096 [7], "Human Factors (HF); Phone Based Interfaces (PBI), Human factors guidelines for the design of minimum phone based user interfaces to computer services". As its title implies, this guideline is not restricted to voice mail or stored voice applications, but may be applied to any phone based interface. It thus includes banking, information retrieval and home shopping applications. Again, no conformance testing procedures are discussed.

Comparison of these documents reveals a surprising lack of consistency. For instance, although each guideline recommends the use of 0 for HELP, the number 9 may be recommended for DISCONNECT, or FORWARD (NEXT). The reason for this may lie partly in the different contexts for which the documents have been produced. France Telecom's (CNET) recommendations are mainly for stored voice services. The ETSI ETR is aimed more generally at Phone Based Interfaces (PBIs), and the VMUIF specification [39] was specific to voice messaging. The ISO document [8] was intended to give generic guidance, as a basic platform for assistance in the development of specific services.

No reference has been found dealing with standardization issues in automatic speech recognition when applied to stored voice services. Since so few applications make wide use of this technology so far, it may be appropriate to consider this as a topic for a future standard or guideline. In this way future developments may be guided towards common minimum levels of interface design and usability. Specific aspects might include coding and concatenation, output speech quality and access codes for languages. Some guidance could also be provided on the extent and minimum content for announcements and explanatory audio messages, and on pre-circulated information to be made available to service users.

ETSI is also in the course of publishing ETR 116 [41] "Human factors guidelines for ISDN terminal equipment design". This will not only contain general and specific human factors guidelines for the design of all user interfaces, many of which are common to other work in related fields of information technology and product design, but also some guidance on the general aspects and methodology for the design process itself. It also emphasises the need for prototyping and user testing during design, as an important aspect of ensuring usability.

It has not been possible to review all sources of standards material for those which may be specific to stored voice services, but it is clear that much material exists to enable designers of such systems, or dedicated terminals, to be able to produce user friendly interfaces in the future. However, some harmonization of recommendations seems desirable. In addition to this specific material, a great deal of generic material exists in the form of more conventional human factors design guidelines and in the literature on human-computer interaction.

What is also clear is that the sample of the telecommunications industry polled seems not to have much knowledge about this material generally, so that ways of raising the general level of awareness needs to be developed. The preparation of general guidelines, or an interface implementation guide dealing with the human factors aspects of stored voice services would provide both the general and some more specific help to service developers and providers.

In order to raise awareness of and to give better information on the design of user interfaces it is recommended that:

Recommendation 8

A suitable guidance document should be prepared for the design of the user interface for stored voice services.

10 The standardization position

10.1 Existing standards

There are a vast number of existing standards which apply to the various types of equipment described in Clause 6 and illustrated in figure 2. Some would be ETSI standards but there would also be many others. For network connection, each terminal will have an appropriate access standard according to the network to which it is connected and the point at which the connection occurs. For example, a stored voice server connected to the PSTN at interface 10 would be expected to meet ETS 300 001 [42] for its access requirements and probably ETS 300 114 (NET 20) [43], for the basic requirements for modems. In addition, another modem ETS or NET may apply together with safety and EMC standards. Similarly a server connected at interface 6 would be expected to meet the appropriate ISDN standards.

The only telecommunications standards particular to stored voice equipment are the interchange protocols referred to in Clause 7 and the user protocols referred to in Clause 8. As far as interchange protocol standards are concerned, the replies to the questionnaire only showed significant usage of the American AMIS protocols [13] and [14] (both analogue and digital), although it is known that many manufacturers use proprietary protocols which offer a greater range of facilities. Even so, most of these are either compatible with, or revert to AMIS protocols when necessary, to give a reduced common set of facilities.

The answers to the questionnaire refer to few recommendations dealing with user protocols probably due to lack of knowledge of the subject by the responder. CCITT Recommendation E.161 [35] defines the MFPB key allocation and the VMUIF specification [39] defines common elements of the service from the user's point of view. Some manufacturers made reference to a guide written by France Telecom [36]. In general, proprietary protocols are used.

10.2 Compatibility of existing standards

To ensure satisfactory end-to-end interworking between voice servers it is necessary to use interchange protocols which are compatible at the higher OSI levels. At present this is being achieved by the widespread use of protocols, mainly developed in America, which at least at the fallback level, are compliant to what have become world-wide industry standards. There seems to be good co-operation between the various suppliers and user organisations to maintain this state of affairs as new functional standards become necessary.

None of the user protocol recommendations that exist at present have any real measure of compatibility. This means that the meaning of any key differs from voice server to voice server and from application to application. There would be clear advantages both to the user and to the equipment manufacturer arising from the adoption of a common set or core set of user procedures for the command and control keys (such as start, stop forward and back) which need not impact on the product differentiation arising from the availability of more complex instructions.

11 Conclusions and recommendations

11.1 Conclusions

In studying replies to the questionnaire it has been found that there was much confusion in the naming and description of the various stored voice services on offer. In spite of this, stored voice services are an active and growing market.

Their are a number of voluntary standards writing fora, made up of manufacturers and users, which are working to ensure that protocol standards are available in advance of the development of the market. As a result, the growth of the stored voice services market is not being inhibited by lack of such standards and so no recommendation is made for new work in this field within ETSI.

The 7% of addressees who responded to the questionnaire appear to lack knowledge of currently existing standards for interworking protocols and user procedures. However it is known that some of the respondents who did not refer to use of these standards, do in fact sell products that actually make use of them.

It is clear that the usability of any given stored voice service has a significant effect on its potential for failure or success in the marketplace. There is scope for ETSI work in this field and a number of recommendations are made which should give improvements in usability and assist in the growth of the market. The convenience of use can often be enhanced by interaction with other messaging services.

Coming new techniques can be expected to have an effect on the manner in which voice services are presented to the customer and on the development of the market. It can be expected that the display facilities available on multimedia terminals will tend to replace some of the provision of information in stored voice form. On the other hand, stored voice messaging may well be considered a useful adjunct to displays, and can give useful assistance to the visually handicapped. These developments may be inhibited in some areas by the non-uniform distribution of supporting communications facilities (such as the availability of DTMF signalling and ISDN access) throughout the European marketplace.

11.2 Recommendations

Recommendation 1

To avoid confusion in the description and specification of stored voice services:

The use of the term stored voice services should be encouraged and the term Audiotex be deprecated in the field of standardization.

Recommendation 2

In order to encourage Companies in Europe to take part in voluntary standards writing fora:

Support should be given to participation in voluntary standards fora that are relevant to stored voice services.

Recommendation 3

In order to assist purchasing decisions in a competitive market:

Methods should be developed to quantify the usability of stored voice services.

Recommendation 4

To assist the growth of a free market in stored voice services it is essential to improve the consistency in their operating procedures:

Guidelines should be developed for the user procedures and dialogue in stored voice services, especially that for entering alphabetic characters on telephone sets with alphabet letters assigned to numbers.

Recommendation 5

To improve the uptake in the use of stored voice services:

Guidelines should be developed for the preparation of introductory and user procedural text and voice messages for use in stored voice services.

Recommendation 6

To assist manufacturers to make more effective use of automatic voice recognition:

Support should be given for the preparation of guidance on automatic speech recognition for stored voice services.

Recommendation 7

To encourage the provision of multi-lingual stored voice services:

A standard for language identification codes should be developed.

Recommendation 8

To raise awareness of and to give better information on the design of user interfaces:

A suitable guidance document should be prepared for the design of the user interface for stored voice services.

Annex A (informative): Bibliography

For the purposes of this ETR, the following documents were used for background information:

1) Communications Week (16 August 1993), p.23-24, CALICCHIO, Dominick: "Voice messaging reaching new users".

2) CCITT Recommendation E.182 (1988): "Application of tones and recorded announcements in telephone services".

3) CCITT Recommendation E.131 (1988): "Subscriber control procedures for supplementary telephone services".

4) News & Views (June 1993), p. 26, COOKE, Jeremy: "Voice processing - speaks for itself?".

5) France Telecom, CNET Lannion A, DEVAUCHELLE, Philippe, POULAIN, Gérard: "General and particular comments about the ISO/IEC/JTC 18/WG9".

6) AFTEL (December 1992), DEVAUCHELLE, Phillippe: "Applications vocales, objectif: qualité et harmonisation ergonomiques".

7) Publication of GTE Laboratories Incorporated, FAY, David: "Interfaces to automated telephone services: do users prefer touchtone or automatic speech recognition?".

8) France Telecom, CNET (13 October 1993): "Standpoint about the ISO/IEC JTC 1/SC 18 N4227 document".

9) Proceedings of the Human Factors Society, 36th annual meeting (1992), p.222-226, GARDNER-BONNEAU, Daryle Jean: "Human factors problems in interactive voice response (IVR). Applications: do we need a guideline/standard?".

10) L'Echo des Recherches, 2e trimestre (1993), n° 152, GUILLOU, L.C.: "Quinze ans de garantie au CCETT".

11) ITU Draft recommendation P.8S (March 1993): "Subjective performance assessment of the quality of speech voice output devices.

12) Computer Communication Technologies for the 90's, ICCC (1988), p.391-396, KOWALSKI, Bernd, WOLFENSTETTER, Klaus-Dieter: "Security for electronic mail and telematic services".

13) News & Views (June 1993), p.24-25, LLEWELLYN, Terry: "A history of voice mail from a UK perspective".

14) IEEE Communications Magazine (April 1993), p.54-60, MASANOBU FUJIOKA et al: "Globalizing IN for the New Age".

15) Communications International (August 1993): "Micro voice target".

16) ISSLS, Paper n° 136", MONESTIEZ, J.C. et al: "Automatization of public fax services".

17) ICCC (1990), New-Delhi, PAWLITA, Peter F: "Computer Supported Telephony: concept of a possible future mass application field".

18) Communications International (February 1993): "PC opens interactive voice to small firms".

19) ISSLS, Paper n° 201 (1993), POULAIN, Gérard: "Voice services: principles, quality and dialogue harmonisation".

20) Byte (November 1993), p.107-108, REINHARDT, Andy: "Partners seek to unite phone and PC".

21) Communications Technology International (1993), p.93-95: "Rotary dial detection for the nineties".

22) France Telecom, CNET Lannion A, SOUBIGOU, André: "Speech recognition and synthesis for ineractive voice services, technologies and human factors aspects. The french experience".

23) Communications International (April 1993): "Talking machines".

24) Teleconnect and Inbound/Outbound magazines (1990): "The voice processing glossary".

25) Security and protection in information systems, IFIP (1989), TROY, Eugene F: "Addressing the telephone intrusion threat".

26) Communications International (August 1993): "UPT: can the phone take the strain?".

27) Proceedings of the Human Factors Society, 36th annual meeting (1992), p.211-215, VIRZI, Robert A. et al: "Skip and scan telephone menus: user performance as a function of experience".

28) Communications International (January 1993): "VM gets the message".

29) Glossary of voice technology terms, Draft n° 1, (August 1990): "Voice Mail Association".

30) Communications International, (November 1993), p.46-51: "Voice processing gets its open hearing".

31) ICCC (1990), New-Delhi, WALSH, Christopher: "Caller identification: social and technical issues".

Annex B (informative): Questionnaire sent to ETSI members and others

European Telecommunications
Standards Institute

ETSI

Request for information about Voice services

You are invited to complete the following questionnaire as carefully as possible. Your answers will be very helpful in identifying any requirements for future possible standardization in this area.

Please send it back to

ETSI
PT36V Leader - Route des Lucioles - Espace Beethoven
F-06921 SOPHIA ANTIPOLIS CEDEX
FRANCE

About your organisation

Company name ______________________________

Contact person ______________________________

Field of activity ______________________________

Telephone number ______________________________

Facsimile number ______________________________

1 General

1.1 Are you acting as :

Application developer	☐
Service provider	☐
Equipment and systems manufacturer	☐
User association	☐
Consulting company	☐
Others *(Please describe your activity below)*	☐

____

____

____

1.2 In which kind of stored voice service applications are you involved ?

Telephone answering	
Voicemail	
Database interrogation	
Database input	
Transactions e.g. Phone banking	
Auto attendant	
Others, please specify	

1.3 Are you an active participant in the Standardization process of voice services ?

In a user association	
Which one?	
In a manufacturer forum,	
Which one?	
In an official Standardization organisation,	
Which one?	

1.4 What do you understand to be the meaning of "Audiotex"

2.1 Terminal Aspects

- From which kind of Terminal Equipment can the voice service be invoked ?

Decadic telephone set	
MFPB telephone set	
ISDN terminal	
Mobile telephone	
Others, please specify	

2.2 Interactive signalling media

2.2.1 Speech band

MFPB	
Voice activation	

2.2.2 Through other channels - please specify-

3 Voice Server Aspects

3.1 General

a) Is your server integrated in a network ?	
b) Is it stand alone Terminal Equipment attached to a network ?	
c) Does your server use a special network access?	
d) Is it part of an Intelligent Network architecture?	

If "c", please explain

3.2 Network interfaces

What network access(es) does your server use?

3.2.1 Public Network interfaces

Analogue interfaces

Loop Start	
2/4 wires E&M	
DDI	
DID	
Others, please specify	

Digital interfaces

ISDN basic	
ISDN primary access	
With DDI	
PCM E&M	
Others, please specify	

If necessary give examples

3.2.2 Special Network interfaces

LAN Ethernet	
LAN Token ring	
WAN	

If necessary give examples

3.3 Inter-Register Signalling

What type of inter-register signalling does your server use

DTMF	
Pulse	
MF-R1	
MFC-R 2	
Others	

If necessary give examples

3.4 Common Channel Signalling

What type of common-channel signalling does your server use?

SS7	
DPNSS	
Others	

If necessary give examples

3.5 Audio Messaging Interchange Protocols

Indicate the message protocol that you use

AMIS-Analog	
AMIS-Digital	
Others	

If necessary give examples

3.6 Links to and interaction with other services

Tick which other services your service links to or interacts with

Facsimile, G3	
Facsimile, G4	
Paging	
Videotex	
TELEX	
Other services	

If necessary give examples

4 Charging and billing aspects

Do you use a premium rate service (kiosk system) for billing ?

Yes
No

If not, how do you recover your fees

If necessary give examples

5 Regulatory aspects

Are voice services subject to regulation in your Country?

Are there any applicable laws

If necessary give examples

6 Human Factors aspects

The following questions about end-user procedures are related to human factors issues.

We are interested in the ease with which an end-user can access and make use of the facilities provided by the application/service/equipment you use or supply.

6.1 Applications description

What is your service called?

How is it described in the:

a)-user guide
b)-abbreviated text on card or other memory aid
c)-menu system on visual display or screen

If necessary give examples

6.2 What controls are used for your service or application?

numeric keys
numeric plus ❋ or #
voice only
voice plus keys
Other

6.3 What kind of dialogue is utilised?

Command	you only enter digit strings by key or voice
Interactive	enter digits, wait for prompt, enter further digits
Mixed	combination of digits, prompts and voice

6.4 What kind of prompts are provided?

tones
voice
text in display window eg.featurephone
text on screen eg. PC

6.7 Is any on-line help provided?

available
not available

6.8 Procedures

Are your procedures based on any published standards or recommendations?

CCITT	
CEPT	
others, please specify	

7 Voice coding and quality aspects

What Voice coding/decoding algorithms do you use?	
What Compression techniques do you use?	
What Transcoding techniques do you use?	
Do you use speech synthesis?	
If yes, what method of synthesis do you use?	
Do you specify any standards of speech quality?	

Annex C (informative): Summary of questionnaire replies

	Comments	Manufacturers	NO & Service	Total
Topic		**dev and R&D**	**providers**	
1.1 Are you acting as :				
Applic developer		14		14
Service provider			7	7
manufacturer		18		18
User association				
Consulting company		3		3
Others	Network operators & Technical support	3	2	5
1.2 In which kind of stored voice service applications are you involved ?				
Telephon answering		20	4	24
Voicemail		17	6	23
Database interrogat		15	3	18
Database input		11	2	13
Transactions		10	1	11
Auto attendant		10	5	15
Others		9		9
1.3 Are you an active participant in the Standardization process of voice services ?				
user association		1	3	4
Which one?	VMAE,'CNET user Ass', AFTEL			
manufactur forum,		3	1	4
Which one?	EHDV, SCSA, MVIP, VMUIF, AMIS			
official organisation		6		6
Which one?	IPAV, CNET, ETSI, CCITT			
1.4 What do you understand to be the meaning of "Audiotex"				
	DefNM3,4,5,6,7,8,9,10,12,13,14 DefM1,4,5,6,9,13,14,15,17,19,21,22	13	5	18
2.1 Terminal Aspects: From which kind of Terminal Equipment can the voice service be invoked ?				
Decadic telephone	Obsolete, no permitted, for paging only	14	4	18
MFPB telephone set		21	8	29
ISDN terminal		14	5	19
Mobile telephone		19	5	24
Others		3		3
2.2 Interactive signalling media, 2.2.1 Speech band				
MFPB		21	7	28
Voice activation		17	4	21
2.2.2 Through other channels - please specify-				
	ISDN, GSM short message, VIDEOTEXT, decadic, outband, voice recognition, data modem	8		8
3 Voice Server Aspects, 3.1 General				
a) in a network ?		15	4	19
b) stand alone T E ?		17	4	21
c) Special access?		5		5
d) Intelligent Netw		10		10
3.2 Network interfaces, what network access(es) does your server use? **3.2.1 Public Network interfaces, analogue interfaces**				
Loop Start		11	3	14
2/4 wires E&M		8	1	9
DDI		13	1	14
DID		7	1	8
Others,	Ground start, PAVI, server in network, MVIP/PEB	3	2	5
3.2.1 Public Network interfaces, digital interfaces				
ISDN basic		7	1	8
ISDN primary		13	4	17
With DDI		9	1	10
PCM E&M		10	2	12
Others	2Mb/s A1,B1assoc., 4bit CAS, Nat prot., SS7, MFCR2, MF socotel, PCM, MVIP/PEB, T1	8	2	10
3.2.2 Special Network interfaces				
LAN Ethernet		13	3	16
LAN Token ring		5	1	6
WAN		6	1	7

	Comments	Manufacturers	NO & Service	Total
Topic		**dev and R&D**	**providers**	
3.3 Inter-Register Signalling				
DTMF		14	3	17
Pulse		10	1	11
MF-R1		3		3
MFC-R 2		11	2	13
Others	SS7, Internal interface,, MFCE, MF socotel, ISDN T2	3	1	4
3.4 Common Channel Signalling				
SS7		8	2	10
DPNSS		4		4
Others	ISDN - D protocol, VN3, 1TR6, SS7 variants	3		3
3.5 Audio Messaging Interchange Protocols				
AMIS-Analog		7	1	8
AMIS-Digital		6		6
Others	CNET Recommendations, X.440, proprietary	4	2	6
3.6 Links to and interaction with other services				
Facsimile, G3		16	2	18
Facsimile, G4	Not yet available	4		4
Paging		15	5	20
Videotex		9	1	10
TELEX		3		3
Other services	Intelligent Network, Subscriber services, E-MAIL, GSM, Tel Services, Modem, msg waiting ind.	9		9
4 Do you use a premium rate service (kiosk system) for billing ? (Y / N)				
		13	5	18
5 Regulatory aspects. are voice services subject to regulation in your Country? YES				
	CNIL, PTT, Common, Tel law, DSP, ICSTIS, DRG	13	5	18
Are there any applicable laws ? YES				
		6	2	8
6.1 Applications description, What is your service called?				
	Many services			
How is it described in the:				
a)-user guide		8	4	12
b)-abbreviated text		9	3	12
c)-On line menu		3	1	4
6.2 What controls are used for your service or application?				
numeric keys		10	3	13
numeric plus ✱ or #		15	6	21
voice only		4		4
voice plus keys		10	2	12
Other	Hook flash, Speech, pulse, Videotex, decadic, Noise/Silent	7		7
6.3 What kind of dialogue is utilised?				
Command by key or voice		5	2	7
Interactive wait for prompt,		7	5	12
Mixed prompts and voice		14	2	16
6.4 What kind of prompts are provided?				
tones		13	4	17
voice		17	7	24
text in display , Fph		5		5
text on screen , PC	MINITEL	5		5
6.7 Is any on-line help provided? (available / not available)				
	Available	14	4	18
	Not Available	3	2	5
Are your procedures based on any published standards or recommendations?				
CCITT	CCITT Rec.E.183	9	2	11
CEPT		3		3
others,	EHDV, FT Rec., Internal Company STD, CNET, Belgacom STD, ISO 1219, BT's in house, EHDV, Proprietary, etc.	9	2	11

	Comments	Manufacturers	NO & Service	Total
Topic		**dev and R&D**	**providers**	
7 Voice coding and quality aspects				
Voice cod/decod algorithms ?	**ADPCM, PCM A law Mu law, DPCM, G711, G721, SBC, proprietary**	**19**	**4**	**23**
Compression techni	**ADPCM, silence, PCM / ADPCM, DSPG, G721, proprietary, SBC, CELP**	**11**	**2**	**13**
Transcoding techniq	**PCM to ADPCM, A law to Mu law, FERMA, SBC, proprietary**	**4**		**4**
Speech synthesis?		**6**	**1**	**7**
Synthesis method?	**CNET Vox, PSOLA (CNET Patent), Proprietary, diphones, TTS, Berkeley speech**	**3**		**3**
Speech quality standards? YES		**1**	**1**	**2**

History

Document history	
June 1994	First Edition
March 1996	Converted into Adobe Acrobat Portable Document Format (PDF)