Speech synthesis in the Intelligent Personal Communication Support System (IPCSS)

Tom Pfeifer

Technical University of Berlin

Open Communication Systems (OKS)

Franklinstr. 28, 10587 Berlin, Germany

e-mail: pfeifer@fokus.gmd.de

Abstract: The Intelligent Personal Communication Support System is introduced as an application for Text-to-Speech (TTS) systems as one of multiple media conversion tools. After discussing intelligent media conversion for terminal flexibility, the necessity of an integrated framework of flexible converters is derived. Evaluation criteria for integrating commercially available TTS systems are explained.

1   Introducing the Personal Communication Support System

In a world of growing mobility in telecommunication and computing it is necessary to integrate the diversifying usage of end-systems and mediums. The user who becomes suddenly reachable everywhere needs control over who may reach him in a specific situation. For consistency of communication the integration of different services would be desirable and can be combined with maximum flexibility of the devices used for access. Speech services are of particular interest because the terminal type 'telephone' can be expected anywhere on this planet, even if other media than audio are not accessible.

Therefore, the Personal Communication Support System (PCSS) is being developed by a team at the institute for Open Communication Systems at the Technical University of Berlin (Dept. of Computer Science) in the context of a BERKOM project for 'IN/TMN Integration'. [1] It is based on the thesis that integrated concepts of Intelligent Networks (IN) and Telecommunication Management Networks (TMN) can be used to provide IN-like control capabilities in the area of service personalization and support of personal mobility.

The PCSS aims to provide the user with functionality like sophisticated call-control, customer profile management, etc. within the emerging distributed multimedia computing and telecommunication environment. It is based on the main principle of integrating the user specific control data for communication services within a logically centralized user profile (though it might be physically distributed), in order to obtain permanent consistency of the data. The main components of the PCSS are shown in Figure 1.

The design supports the user's control of his reachability (at maximum or filtered) – independent of his location, the used communication media, and the applied communication mode (asynchronous or synchronous), as well as the organization of communications (when, how and with whom the user communicates, as the originator and as the recipient).

One of the most prominent principles of the PCSS, distinguishing this system from traditional telecommunication approaches, is the mostly *person*-oriented and *location*-oriented operation mode - in contrast to *number*/network-address-oriented modes of generic UPT (Universal Personal Telecommunication) systems. [2]

Currently realized components of the PCSS encompass a User Profile Management Service, a

Visitor Information Service, and a Profile Administration Tool as user front-ends, all of them interacting with the core PCSS, which implements a Generic Service User Profile. The latter contains all information required to support the personal mobility, the personalization of communication services, and the advanced user information services.

The rules to handle communication requests are derived from static and dynamic information provided for each user, that is stored in his profile consisting of directory entries and managed objects. In compliance with telecommunication standards, static information like default phone numbers, mail addresses, regularly activities are hold in X.500 directories [4], while dynamic information like current location in the building, call logic information, pointers to prerecorded messages etc. are hold in X.700 [5] managed objects.
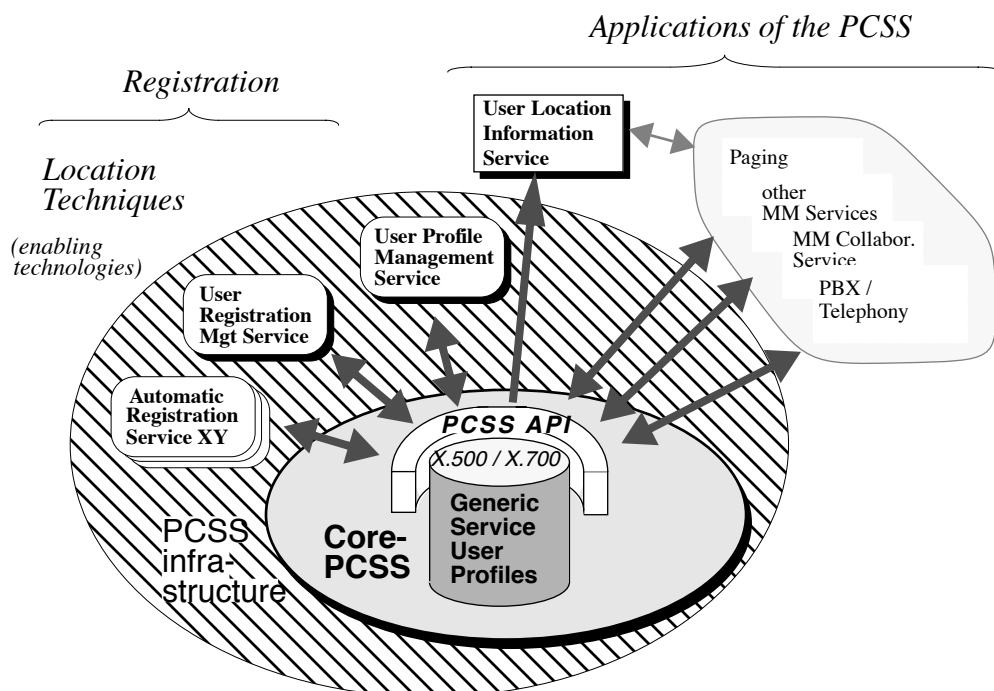

Figure 1: PCSS: functional overview [2]

Location information (i.e. the room or zone a user currently stays in) is derived from manual registration (login, card readers), predefined personal schedules, as well as from automatic registration by wearing electronic badges [3]. If the PCSS is used for call handling, a four-step mapping would consider reachability,

2   Intelligent PCSS

The project of the Intelligent PCSS (IPCSS) focuses on the extension of the existing functionality regarding service personalization and service integration. The main aspects of this development are

- Support of *Terminal Flexibility*, i.e. the dynamic selection of an appropriate communication terminal system, considering:
  a) the availability of systems currently in reach of the user, and
  b) the suitability for the requirements of the chosen form of communication.
  The ability to *convert* different kinds of media into each other is one of the major tools for the realization of this aspect.
  One very typical example is the conversion of text into speech, if the available terminal for

the perception of received e-mail is a public telephone.

- *Service Integration*, i.e.the possibility to define individual, 'intelligent' service interworking scenarios by each user. He shall be able to define the dependability between his means of communication in his *Call Logic Matrix*, e.g. the forwarding of his e-mails to a certain fax machine, the storage of incoming phone messages in his voice mail box and forwarding as multimedia mail. Converting different kinds of media is required by this aspect also.
- Extension of realized PCSS applications for support of the Service Integration.

3   Related work

Communication supportive systems emerging currently on the market are mostly insular solutions, limited to a fixed corporative environment, focusing on either enhancing telephone services or providing dedicated video conferencing utilities. Few approaches exist so far to integrate the varieties of communication modes in both, the telecommunication and the computing area.

A network architecture with distributed user-agents has been proposed by the Fujitsu Laboratories [6, 8]. It aims towards a transparent communication from the user's point of view, and easy methods for customizing personal services. In the proposal a personal ID is employed which is transparent through all kinds of networks.

Schmandt [7] states that mobile 'nomadic' computing services could be built with currently available hardware. He describes an experimental environment at the MIT which tries to integrate – at the service and user-interface levels – the multiple media into a cohesive nomadic information infrastructure with a graceful transition from desktop to nomadic locales.

A proposal by the IBM research center [9], drafting Smart Networks, emphasizes the goal of Best Method Delivery, depending on the equipment in reach of the user. Among the tools for obtaining this goal are inferencing systems, mobile agents, media transformers, and supportive gateways.

A team in the RACE project [10, 11] studies a personal communication environment resembling the PCSS in some components, although it is focused on synchronous communication requests. However, the chosen approach lacks the intelligent media conversion as well as the automatized location techniques. In further work [12], the concept of service translations (conversions) has been added to the project.
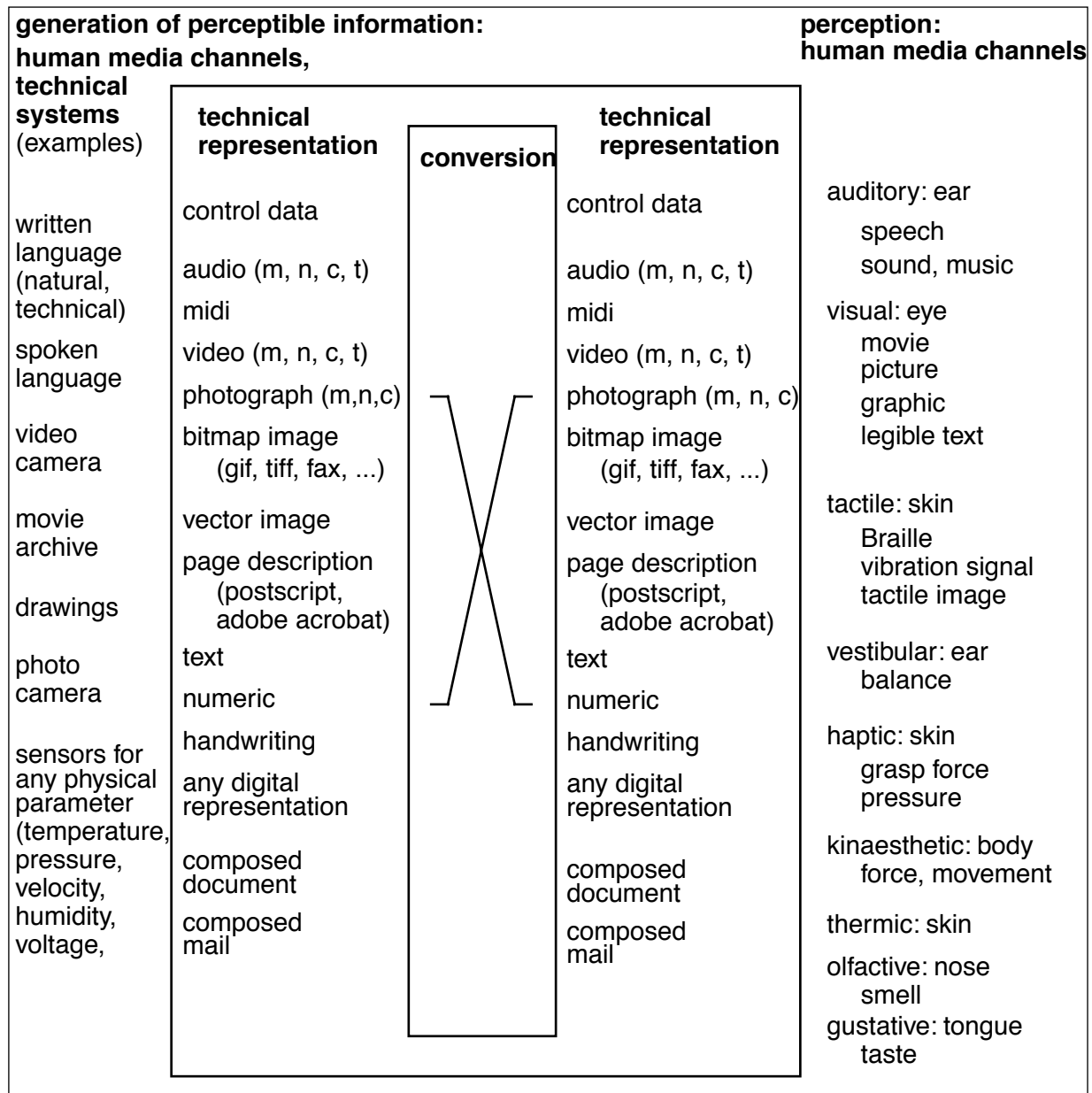
In the context of integrating IN, TMN and distributed computing the most promising candidate is the Telecommunication Information Networking Architecture (TINA) [21]. Its software architecture for information networking is based on distributed computing technologies, object-oriented design methodologies, network and service management an IN principles, with a service driven approach.

Regarding TTS solutions most manufacturers offer text reading and e-mail vocalization as one of their primary applications. However, in most cases the service is not generic but proprietary, tailored to a dedicated environment of operating system and hardware platform. Mostly, audio is sent to the speaker line without the possibility of further processing, audio format conversion and forwarding.

4   Intelligent Media Conversion for Terminal flexibility

For supporting terminal flexibility in the discussed environment lots of different conversions

must be possible. They need to have unified interfaces and must be capable of combination. Therefore, generic solutions in a framework are required instead of isolated, proprietary, special-purpose ones.

generation of perceptible information: human media channels, technical systems (examples)	technical representation	conversion	technical representation	perception: human media channels
	control data		control data	auditory: ear
written language (natural, technical)	audio (m, n, c, t)		audio (m, n, c, t)	speech sound, music
	midi		midi	visual: eye
spoken language	video (m, n, c, t)		video (m, n, c, t)	movie picture
	photograph (m,n,c)		photograph (m, n, c)	graphic
video camera	bitmap image (gif, tiff, fax, ...)		bitmap image (gif, tiff, fax, ...)	legible text
movie archive	vector image		vector image	tactile: skin

	page description (postscript, adobe acrobat)		page description (postscript, adobe acrobat)	Braille
				  vibration signal
				  tactile image

drawings	text		text	vestibular: ear
photo camera	numeric		numeric	balance
	handwriting		handwriting	haptic: skin
sensors for any physical parameter (temperature, pressure, velocity, humidity, voltage,	any digital representation		any digital representation	grasp force pressure
	composed document		composed document	kinaesthetic: body force, movement

	composed mail		composed mail	thermic: skin
				olfactive: nose
				  smell
				gustative: tongue
				  taste

parameters:
m, n: media dependent parameters
    (frame/sampling rate, quantization, resolution, size, color depth, etc.)
c: applied compression technique
t: time, duration, etc.

Figure 2: Generic conversion matrix

Beyond support of terminal flexibility are other reasons for media conversion like:
- unavailability of certain media
- harmonization of different format requirements of the partners in communication (e.g. different audio representations like μlaw and Sun-audio)
- volume/bitrate reduction for transmission and storage
- integration of disabled (bodily challenged) people into communication to compensate the

loss of a sense (e.g. Braille, speech synthesis, OCR, tactile images for blind people)

As a first step in the IPCSS, conversions are required for e-mail to speech (TTS), announcements to speech (TTS), fax to mail (OCR), fax to speech (OCR+TTS) and mail to fax. In the area of speech recognition first applications are command recognition (from phone terminal) and speaker recognition (for prioritizing and security), as well as voice-mail recognition at a later stage.

5   Integrated framework of flexible media converters

It is necessary to define a generic approach to media conversion in order to obtain
- a maximum of flexibility
- replacability and modularity of components
- common interfaces for management, control etc.
- an object-oriented design
- a way to integrate commercial solutions respecting the interests of the manufacturer.

The whole set of theoretically possible media conversions can be shown in a matrix like Figure 2. While it is theoretically demanding to convert every possible medium into any possible other, practical considerations will lead to discretions that are required in a given environment.

Table 1 gives a collection of practical examples for illustration, focusing on requirements in

source	drain	process	example / application
bitmap image	bitmap image	format conversions	tiff ->JPEG
video	video		MPEG -> H.261
audio	audio		µ-law -> a-law
bitmap image	vector image	vectorization	
audio	vector image	visualization	length of message
text	speech	speech synthesis	TTS
speech (audio)	text	speech analysis & recognition	commands, dictation
bitmap image	text	OCR	OCR
fax bitmap	speech	OCR+TTS	fax reading
temperature	bitmap image	temperature distribution of objects	weather map, medical map
temperature	audio	TTS: value reading	weather report
numerics	image	visualization of statistics	charts
text	tactile information	feed Braille output device	blind reading
control data	tactile information	vibration device	pager signalling
audio	control data	speaker recognition	prioritizing, authorization
photograph or video	text	face recognition, mimic recognition	(e.g. very low bitrate compression)

Table 1: Selected examples of media conversion

multimedia communication, with the attitude that even conversions that sound strange in the first place might be of very practical relevance.

The range of conversions varies tremendously in effort, cost and required resources. Some kinds are easy to implement with two lines of C code without any patent or license require-

ments, while other are highly complex, requiring sophisticated solutions that might be available only as commercial products, and may be covered by patent or copyright protections. Some conversions are purely algorithmic, while other require approaches of artificial intelligence, or require pipelined processes of decoding, editing, and re-encoding.

A way that is often used to reduce the number of required tools in conversion is the use of a generic intermediate format. The latter may be of the same kind than the converted medium (e.g. image format conversion), or a different one (e.g. fax > (ocr) > text > (speech synth.) > reading). In order to avoid losses in quality the intermediate format needs the resources of the highest possible quality among the involved formats.
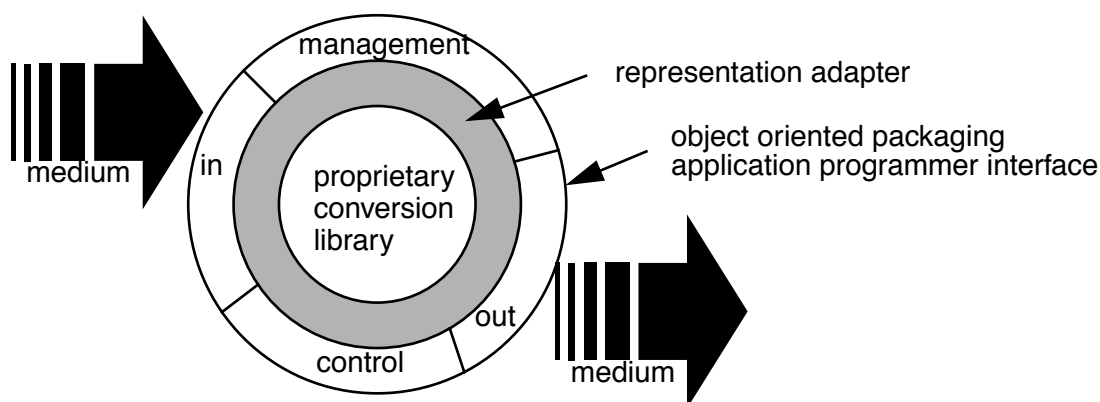
The advantage of requiring less tools is paid with
• the necessity of more resources (e.g. for an intermediate format or for computing time)
• the necessity to convert twice, causing a delay and – in some cases – a reduction in quality

A possible solution for this problem is a hybrid approach, i.e. to use dedicated tools if available, and intermediate formats in other cases. At this point, it becomes necessary to classify the demands of conversion:

1. highest priority, very often used, maximum speed required:
   This category should be implemented as a dedicated service, featuring a dedicated software solution including supplemental hardware, for performing a one-step conversion.
2. medium priority, often used:
   In the intermediate area compromises are possible, dedicated software would be an option, one-step conversion recommended.
3. lower priority, rarely used:
   For services that are seldom employed the conversion via a generic intermediate format is possible. The resulting multiple-step conversion requires a preliminary definition of a conversion path, e.g. instructions which converters and intermediate formats are appropriate to achieve the maximum quality or shortest conversion time, respectively.

Even when the same pair of media is converted, different performance, quality and gradability might be required – depending on the application. E.g., speech recognition (speech –> text) can be focused on dedicated speakers for dictation purposes, or on a very limited vocabulary produced by unknown speakers for command or keyword recognition.

Figure 3: Generic converter model

A generic converter model is shown in Figure 3. A representation adapter layer covers manufacturer specific properties of the core library. The outer interface has to realize the object ori-

ented behaviour as well as the provision of management and control interfaces. The model is designed to comply with the requirements of Computational Objects in TINA [19].

6   Evaluation criteria for available TTS systems

A list of possible TTS solutions that are ready to use (see Table 2) has been derived from various sources and is currently under evaluation [16, 17, 18]. For application in the designed system, a couple of further criteria has to be considered beyond a respectable quality of the produced speech. Most important is the usability of an application programmer interface (API) for the integration into the whole system. The availability of source code is not requested, as it would violate the commercial interests of the manufacturer.

A cross-platform availability is a desirable property in a heterogeneous environment, though it is not mandatory, because a dedicated TTS server in a department or corporation may be chosen according to the requirements of the TTS system.

The TTS must produce a generic digital signal as audio output, which might be subject of further processing, format conversion and forwarding as a following step to speech generation.

product	manufacturer	operating system	hardware platform	supplemental hardware	languages
TrueTalk	Entropic	UNIX	Sun, SGI		E,
VisualVoice	Stylus Innovations	DOS/WIN	PC	sound card possible	E,
EASE	Expert Systems			Dialogic required	E,
TrueVoice	Centigram	UNIX, DOS	Sun, PC		E,
Dialogic TTS	Dialogic	DOS	PC	Dialogic required	E,
DecTalk	Digital				E,
Lernout & Houspie	Lernout & Houspie	OS/2, DOS/WIN, etc.	various		E, Ger, F, NL, Esp,...
Rhetorex TTS	Rhetorex	UNIX, OS/2, NT	PC	Rhetorex required	E,
BestSpeech	Berkeley Speech-Technologies	DOS, OS/2, UNIX	PC etc.	Dialogic, Rhetorex, etc. possible	E, Ger, F, I, J, NL, Rus
Infovox	Telia Promotor Infovox	DOS	PC	Infovox 500	Ger
Elan Informatique	ELAN Informatique	DOS	PC	Televox Psola-8m	Ger, F

Table 2: Evaluation of available TTS systems

7   Future work

Instead of being statically implemented in a service environment, an interworking of converters with or as Intelligent Agents [13, 14, 15] might be useful. By the definition provided in [15], a converter would work as an Intermediary Agent, "providing common services needed by other agents".

Provided with the required converters, the set of terminals inside a room or zone can be defined as a virtual access point profile, enabling the dynamic selection of terminals for current communication requests.

As a major development the integration of the IPCSS into a TINA platform is planned [19]. Therefore, existing and new components of the IPCSS – including converters – have to be matched with the requirements of Computational Objects in TINA [20].

8 Summary

In this paper an application for speech synthesis (TTS) in information systems has been provided. The Intelligent PCSS will include this kind of media conversion as one of multiple conversion tools, equipped with a unified interface for management and control. The requirements for integrating a TTS system into such a framework have been discussed.

9 References

[1] Berkom II Project "IN/TMN Integration", Deliverable 6: "Personal Communication Support System: Preliminary Specification of Profiles and Management Services" - Technical University of Berlin, Inst. for Open Communication Systems (OKS), June 1995

[2] Eckardt, T.; Magedanz, T.; Pfeifer, T.: On the Convergence of Distributed Computing and Telecommunications in the Filed of Personal Communications. - in Proc. of KIVS'95, Franke, K. et al. (Ed.): Kommunikation in Verteilten Systemen, Springer: Feb. 1995, Chemnitz, Germany, pp. 46-60

[3] Harter, A.; Hopper, A.: A Distributed Location System for the Active Office, in: IEEE Network, Special Issue on Distributed Applications for Telecommunications, January 1994

[4] ITU-T Recommendation X.500 / ISO/IEC/IS 9594: Information Processing - Open Systems Interconnection - The Directory, Geneva, 1988

[5] ITU-T Recommendation X.700 / ISO/IEC/IS 7498-4: Information Processing - Open Systems Interconnection - Basic Reference Model - Part 4: Management Framework; Management Framework for CCITT Applications

[6] Iida, Ichiro; Nishigaya, Takashi; Murakami, Koso (Fujitsu Laboratories Ltd., Japan): DUET: Agent-based Personal Communications Network. - in: Proc. of the XV. Intern. Switching Symposium, ISS '95, April 1995, Berlin, Germany, Vol. 1, pp. A2.2: 119-123

[7] Schmandt, Chris: Multimedia Nomadic Services on Today's Hardware. - in IEEE Network, Sept./Oct. 1994, pp. 12-21

[8] Nishigaya, Takashi; Iida, Ichiro (Fujitsu Laboratories): Distributed Communication Control Architecture for Personal Communication Service. - in: Proc. of the 3rd Int. conf. on Universal Personal Communications, ICUPC '94, Sept. 1994, San Diego, California, pp. 272-275

[9] Harrison, Colin G.: Smart Networks and Intelligent Agents. - IBM T. J. Watson Research Center; technical report

[10] Guntermann, M.; et al.: Integration of Advanced Communication Services in the Personal Services Communication Space - A Realisation Study. - in: Proc. of the RACE IS&N Conference 1993 (Mobilise), pp. II/1/p.1-12

[11] Abramovici, Martine; Klußmann, Niels: Graphical User Interface Style Guide for Mobile Communication Services. - in: Proc. of IS&N '94, 2nd Int. Conf. on Intelligence in Broadband Services and Networks, Sept.94, Aachen, Germany, pp.89-97

[12] Niebert, Norbert; Geulen, Eckhard (Ericsson Eurolab): Personal Communications - What is Beyond Radio? - in: Proc. of IS&N '94, 2nd Int. Conf. on Intelligence in Broadband Services and Networks, Sept.94, Aachen, Germany, pp. 247-257

[13] Rizzo, Mike; Utting, Ian A.: An Agent-based Model for the Provision of Advanced Telecommunications Services. - in: Proc. of TINA'95, Integrating Telecommunications and Distributed Computing – from Concept to Reality, Feb. 1995, Melbourne, Australia, pp. 205-218

[14] Reinhardt, Andy: The Network with Smarts. New agent-based WANs presage the future of connected computing. - in: Byte Oct., 1994 pp.51-64

[15] Atkinson, Betty; Brady, Stephen; Gilbert, Don; et al. (IBM Corporation): IBM Intelligent Agents. - in: technical report, pp. 265-277

[16] Jainschigg, John: Text-to-Speech. Sobering up the "drunken swede". - in: Teleconnect, May 1995, pp. 125-128

[17] Berkeley's Bestspeech Text-to-Speech. - in Teleconnect, May 1995, p. 135

[18] Léwy, Nicolas; Hornstein, Thomas: - Text-to-Speech Technology. A Survey of German Speech Synthesis Systems. - UIBLAB Technical report 94.10.2, ftp://ftp.uiblab.ubs.ch/pub/paper/GermTTS.ps.Z

[19] Eckardt, T.; Magedanz, T.; Popescu-Zeletin, R.: Application of X.500 and X.700 Standards for Supporting Personal Communication in Distributed Computing Environments. - in: Proc. of the 5th IEEE Computer Society Workshop on Future Trends of Distributed Computing Systems, Aug. 1995, Cheju Island, Korea, pp. 232-241

[20] Telecommunication Information Network Architecture – TINA-Consortium: Service Architecture. March 1995

[21] Dupuy, Fabrice; Nilsson, Gunnar; Inoue, Yuji: The TINA Consortium: Towards Networking Telecommunications Information Services. - in: Proc. of the XV. Intern. Switching Symposium, ISS '95, April 1995, Berlin, Germany, Vol. 2 B5.2 pp. 207-211