ESCA - NATO/RSG 10 Workshop on
Applications of Speech Technology
Lautrach, Germany
September 16-17, 1993

INTEGRATED APPLICATION OF LOW BIT RATE CODING AND TEXT-TO-SPEECH SYNTHESIS FOR INTELLIGENT TELECOMMUNICATIONS SERVICES

Géza Németh, Attila Tihanyi, Gábor Olaszy* and József Gátmezei**

Department of Telecommunications and Telematics, Technical University of Budapest,
Budapest, Sztoczek u. 2. Hungary 1111
Tel.: +36 1 1664 011 ext 24-01, Fax: +36 1 1812 302, E-mail: h1796nem@ella.hu

**Phonetics Laboratory, Linguistics Institute of the Hungarian Academy of Sciences*
Budapest, P.O. Box 19. Hungary 1250
Tel.: +36 1 1557 122 ext. 218, Fax: +36 1 1758 128

***Hungarocom Telecommunication Ltd.*
Budapest, Temesvár u. 20. Hungary 1116
Tel.: +36 1 1869 522, Fax: +36 1 1669 320

ABSTRACT

In the paper speech technology aspects of a voice announcement system (called Intelligent Message System, IMS) designed for providing real-time multichannel services for crossbar and/or for stored program control (SPC) exchanges will be introduced. The most important messages are coded from human voice at a bit rate of 1.5 kbits/s. Additional messages can be created instantly on the spot, from existing vocabulary elements or by using the MULTIVOX multilingual text-to-speech system /1/, enhanced with extra editing and downloading capabilities. An approximately double Eurocard format board contains a Voice Synthesizer Subsystem (VSS), that can handle maximum 16 channels in real-time from a vocabulary of approximately four minutes. 10 VSS boards (i.e. 160 channels) can be controlled together using the multipoint RS-422 interface. The system has been installed in six large exchanges of approximately 160.000 lines between December 1992 and May 1993 .

Keywords: *intelligent telecommunications services, speech synthesis, voice announcement, messaging systems, exchange extensions*

1. INTRODUCTION

In the course of the reconstruction of the telecommunications system of Central- Eastern European countries, one of the greatest problems is the heterogeneous nature of the existing systems. In Hungary e.g. rotary, crossbar and stored program controlled (SPC) digital exchanges from different vendors have to work together. It is very difficult to introduce intelligent services in this situation, because all subscribers await to have access to the same services (at least in the same town). This requirement can be hardly met by the above mentioned technologies.

The situation may become extremely critical if there are large-scale changes in the network. However, in the course of dynamic expansion, the change of the directory number of numerous subscribers is inevitable. The information regarding the changed numbers within a location is still manageable by traditional means (e.g. advertisements, special inquiry numbers, etc.) but in case of changes in connection with putting into service new exchanges all over the country, the provision of information in a traditional way is almost impossible.

Calls toward changed numbers generate excess traffic which can overload exchanges. If only the fact of the change of the directory number is announced, it draws a mass demand for inquiry services, which cannot be easily handled. The application of individual announcement devices per line is uneconomical. There is also a problem of what to do with these devices after the period when the new directory numbers are already evident. There was no straightforward solution to these problems for crossbar exchanges.

To avoid the above mentioned problems the Intelligent Message System (IMS) has been developed. Besides solving the problem of mass inquiry during number changes it assures the future use of the equipment to provide new value added services for the subscribers.

2. SYSTEM DESCRIPTION

The basic subsystems of IMS system development can be seen on *Figure 1.*

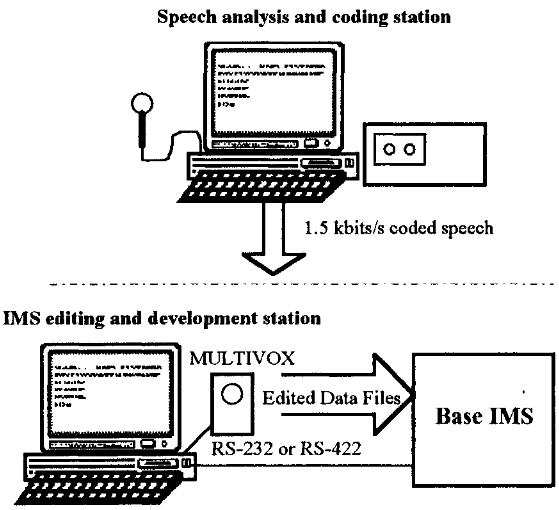

Figure 1. IMS system development configuration

The **Speech analysis and coding station** is used for coding human speech into a 1.5kbits/s data stream required by the formant speech synthesizer chips used in the Voice Synthesizer Subsystem (VSS). It is separated by a dashed line from the other subsystems of the figure as it can be operated by a highly trained speech specialist only.

The **IMS editing and development station** is a PC based system and it is used for compiling the required vocabulary for a VSS application. The vocabulary may consist of maximum 255 ***elements*** at a length of four minutes. Elements may be derived from the output of the speech analysis and coding station or from the codes produced by the MULTIVOX multilingual text-to-speech system. An earlier implementation of this principle is described in /2/. It is possible to manually enhance and edit the data using a character or graphic oriented editor. The elements may be burned into the EPROMs of a VSS or downloaded into RAM. Messages may be either these elements or one of maximum 255 ***announcements***, that are defined as a series of vocabulary elements. After some training the IMS editing and development station can be effectively used by an engineer or technician at a centralized location or at a telephone exchange.

The **base IMS** is a multi-processor real-time data acquisition and processing system, connected to the telephone exchange. A single IMS is capable of fulfilling the voice announcement needs of approximately 40.000 subscribers. From a pre-defined vocabulary it is possible to allocate a message by either the operator or the subscriber to any subscriber line. Maximum eight VSSs (i.e. 128 channels) can be operated within the IMS simultaneously.

The IMS uses the R2-MFC signals of the exchanges to determine the called number. It extracts the corresponding message code from its' database then and re-routes the call to a free channel of a VSS. It sends the message code to the VSS through the control bus (RS-422) concurrently. In case of number changes the announcement starts within less then two seconds.

Two possible IMS configurations in an Ericsson ARF crossbar environment are illustrated on *Figure 2.*

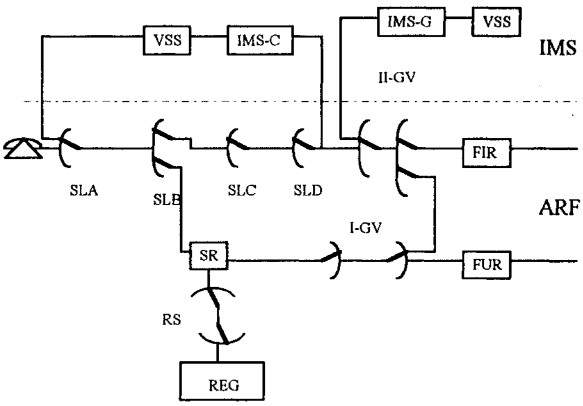

Figure 2. IMS installation possibilities in an ARF crossbar environment

The IMS-G configuration is suggested if complete exchange branches are changed. When the call is connected to such a direction the group selector stage activates the IMS, which requests the last digits of the called number by means of MFC signalling. It then looks up the new number in its' database and sends the corresponding control data to a VSS.

In case of large scale changes within a subscriber stage, the IMS-C configuration can be applied. The IMS is connected to the control section of the subscriber stage and requests the digits of the called party. If it is a number that has changed, the new number is announced by a VSS.

3. VOICE SYNTHESIZER SUBSYSTEM (VSS)

3.1. VSS hardware architecture

The VSS is an intelligent standalone speech output peripheral implemented on a single, approximately double Eurocard sized printed circuit board (PCB). It requires a single +5V DC power supply and can be controlled through the serial multipoint RS-422 interface by a master device. The VSS is configured to be a slave. Its' address can be set anywhere between decimal 0 and 15. It provides maximum 16 symmetric, overvoltage protected analogue output channels, that can be directly connected to the subscriber line. The layout of the VSS is given on *Figure 3.*

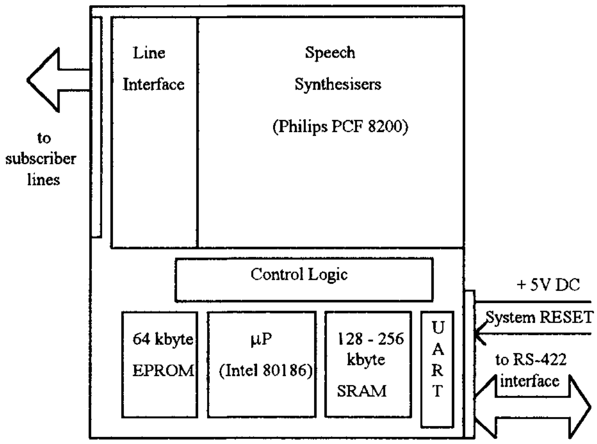

Figure 3. Voice Synthesizer Subsystem (VSS) layout

The VSS's local processor is a 16 bit Intel 80186 microprocessor. Due to its' PC compatibility, flexible memory management and interrupt capabilities, it proved to be an easy to use solution. Only minimal logic was needed for memory, speech synthesizer and UART control. Four speech synthesizer chips were mapped to each interrupt level of the 80186. The 8250 UART was added to the highest priority interrupt. The board contains 64 kbyte EPROM and 128-256 kbyte static RAM memory.

In order to be able to use this relatively simple hardware for handling 16 channels in real-time, a low bit-rate formant synthesizer (1.5kbits/s, Philips PCF 8200) was chosen. Conventional PCM store-and-forward techniques had to be excluded because of two reasons.
i) At 8 kHz sampling rate 125/16 < 8 µs service time/channel would be required,
ii) A four minute vocabulary would demand even in case of 16 kbits/s ADPCM 480 kbyte memory versus the 45 kbytes needed by formant coding.

Because of limitations in the switching control equipment only maximum 15 kbytes can be downloaded into RAM. But this is enough for storing vocabulary elements at a total length of 80 seconds. By careful coding of the vocabulary elements, the quality problems usually associated with such a low bit rate could be practically eliminated.

3.2 VSS services

The VSS provides the following basic services:
- output of messages,
- vocabulary download into RAM,
- announcement definitions download into RAM.

The output of messages can be controlled in several ways, which are briefly introduced below. A single 7 byte command record may refer to the following:

Byte No.	Function
1.	VSS no. (0-15)
2.	Channel no. (0-15)
3.	message type 0: stop speech on channel 1: stop speech on channel, announce number change (NO1) and repeat it twice 2: say a vocabulary element 3: stop speech on channel, say an announcement definition and repeat it twice 4: stop speech on channel and send busy tone to it for 15 seconds 5: stop speech on channel and send ringing tone to it for 15 seconds 6: stop speech on channel and send error tone to it for 15 seconds 7: stop speech on channel, announce regional + directory number change (NO2) and repeat it twice 8: stop speech on channel, announce directory number change once in Hungarian and twice in English (NO3) 9: stop speech on channel, and say vocabulary elements given in Byte4-7 10: stop speech on channel, and say vocabulary elements given in Byte4-7 + the following record twice 11: stop speech on channel, and say vocabulary elements given in Byte4-7 + the following two records twice
4.-7.	If Byte3 = 1, 7 or 8: phone numbers to announce in BCD coding If Byte3 = 2, 9, 10 and 11: vocabulary element codes (1-255, 0 if nothing more to say)

This command structure ensures flexible and extendible definition of messages to be announced. There are presently three modes of announcing the maximum 8 digit telephone numbers given in Byte 4-7. NO1 is intended for local Hungarian announcements, NO2 is suggested for long distance calls, while NO3 is suggested for areas, where foreigners frequently call in (e.g. hotels, business offices, etc.). New options and languages (e.g. German, Italian, Spanish, Dutch, Finnish or Arabic) can be easily added.

The largest VSS configuration installed until the time of writing contained 100 vocabulary elements and 20 announcement definitions, the largest of which had 30 vocabulary elements. It can be favourably compared to the 14 fixed messages that are generally available in SPC exchanges of the Hungarian network.

If new messages have to be added to the EPROM based vocabulary elements for operational or testing purposes, a vocabulary of up to 50 elements at a total length of 80 seconds can be downloaded to the VSS even during operation, too.

It is possible to download to RAM 200 announcement definitions at a length of maximum 100 elements each, as well.

All RAM data are continuously checked by on-the-fly checksum calculation. It also ensures that short RESET pulses or power problems do not necessarily destroy RAM data.

A 300 character (approximately 5 minutes) long message buffer is provided for each channel, so that delay in serial line access should not cause system breakdown or audible pauses of the messages in a channel .

4. CONCLUSIONS

The IMS system has been installed in six large exchanges of approximately 160.000 lines between December 1992 and May 1993 in Hungary. Because of the urgent need of such an intelligent service, no previous acceptability tests could be conducted. Due to the relatively large scale application in a short time some valuable user and operator feedback has already been provided with very favourable results (e.g. 10 user reclamations in one month for 3000 lines).

Quite surprisingly the users very much appreciated the fast response time (less than 2 seconds) of the system. Traditional PCM store-and-forward techniques often couse a delay of more than 20 seconds. Most of the complaints regarded pause timing within or between numbers or other vocabulary elements. Fixing these errors would require regular testing on a larger population and/or the introduction of caller modifiable options.

From the operator side there was a very positive feedback as well. Using the IMS a number change can be handled with a hold time between 20-30 seconds. Manual operators need one more extra call and a hold time of one minute at best, so much more traffic is generated.

As the number change database can be filled up before and activated parallel to the physical connection of cables, the announcement services can be closely coupled (in virtually seconds) to the dynamically changing network environment. Announcement of several thousand numbers can be changed in a few minutes. The similar manual processing took approximately one week.

Although at the time of development providing intelligence for crossbar exchanges was at the primary focus, during the course of application it turned out that the IMS could be well used in SPC exchanges, too. According to our knowledge unfortunately no other system is capable of analyzing all digits of a call by R2-MFC signalling (i.e. providing 8 digit by 8 digit mapping of number changes).

The VSS subsystem also provides much more flexible services and larger vocabulary than any other similar system available. Taking into account its' compact, easy-to-use design it might be used for other announcement (e.g. manufacturing) purposes as well.

REFERENCES

/1/: *Olaszy, G.; Gordos, G.; Németh, G.:* The MULTIVOX multilingual text-to-speech converter. in: G. Bailly, C. Benoit and T. Sawallis (eds.) 'Talking machines: Theories, Models and Applications', Elsevier, 1992, pp. 385-411

/2/: *Németh, G.; Czap, L.; Gordos, G.; Koutny, I.; Olaszy, G.; Tihanyi, A.:* An integrated approach to text-to-speech and fixed vocabulary formant synthesis. Proc. of the VDE-ITG Digital speech processing seminar, Bad Nauheim (Germany), 26-28 October 1988, VDE Verlag, pp. 213-217