Assistive Technologies

Ion SMEUREANU, Narcisa ISĂILĂ
Academy of Economic Studies, Bucharest
smeurean@ase.ro, isaila_narcisa@yahoo.com

A special place into assistive technologies is taken up by the speech recognition and speech synthesizer, which can be used by many different users, persons with visual, language or mobility disabilities. For many years soft developers have been concerned by speech recognition and text-to-speech because we assist to great changes in informatics area and accessibility is the main condition in the creation of assistive software applications.
Keywords *: speech recognition, text-to-speech, assistive technologies, accessibility.*

Introduction

Over a decade, the computer asserts oneself as a needful instrument for persons with disabilities offering them a new perspective, a new way to live.

For these users, the products which have existed on the market offer additional accessibility to computer and are created for each type of disability. For example, assistive technology for visual disability persons includes the following products: screen enlargers, screen readers, screen review utility, speech recognition systems, speech synthesizer, refreshable Braille displays, Braille embossers, text readers, word prediction programs. Assistive products for persons with mobility disabilities include speech recognition systems, editing programs on screen using alternative products (sip and puff, sticks, joysticks, trackballs), alternative keyboards, keyboard filters for editing or touch screens. The persons with learning impairments can use word prediction programs, reading comprehension programs, speech synthesizers, speech recognition programs.

Many accessibility characteristics are offered by Windows operating system or Office package, which render easy the access to different elements using keyboard or mouse. Into operating system it's remarkable the accessibility offered by Magnifier utility (for increasing the zoom of some screen parts), Screen Review (which reads information from screen using the sound), Narrator (for helping the user who works with programs from Control Panel, NotePad, WordPad or Internet Explorer browser). There are a lot of settings in Control Panel using Accessibility Options application regarding contrast, filter options, colors or navigation, all to increase accessibility to each element of application used by the user.

Office package includes accessibility characteristics as: zoom, contrast between graphic elements, the possibility to page setup for better seeing documents' content (using Reading Layout), development accessibility Web pages (Microsoft FrontPage) by adding text to images, formatting styles, making image maps.

1. Speech Recognition Systems

Named also voice recognition programs, may let users to introduce data, using voice instead keyboard or mouse.

The main characteristics of the speech recognition system are:
- vocabulary's size, the number of words recognized
- the separated or continuous speech
- the conditions of noise
- the number of speakers
- the percentage of recognize
- processing time, on-line, delay, off-line
- the area of applicability

The speech recognition is a complex process composed by difficult constructive parts, in which a part of the system, the physics part, converts the sound into electric signal and adjusts it for entrance into the next part. The second part, the logic part, is represented by the computer with sound board and necessary program for all required processes.
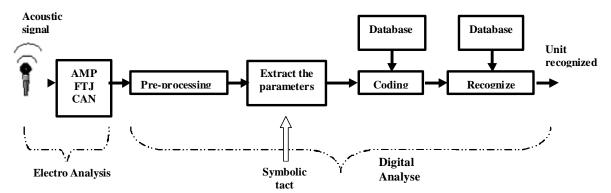
Fig.1. General structure of a speech recognition system

For understanding human speaking, each speech recognition system uses four main components with strong connection:

1) *Dividing text in words*, process used by speech recognition engine, which assures greater or less accuracy in speech recognition. Thus:

a) digitization of the speech, meaning to insert short pauses after each pronounced word, which assures finding by the system the beginning and the end of the word. The advantage is a little power calculation but the system becomes unused if we don't respect pauses between words.

b) identifying words into vocabulary method, which allows the user to speak naturally without pauses between words, but the system can read wrong if some used words aren't into the vocabulary.

c) continuous speaking method, offers the best accuracy in speech recognition processing every pronounced word. Because the system doesn't use elements to separate words is necessary a longer period of time to find the beginning and the end for every word.

2) *The vocabulary or words list*, which speech recognition system can find to a certain moment. Using a rich vocabulary, that means to improve speech recognition but to increase the size vocabulary isn't a guaranty for a better accuracy if many words have the same meaning.

3) *Finding words* means to search in vocal database and to achieve connections to each other and write the audio signal.
There are two methods:

a) identifying the whole word, a method which consists in finding into the database for the word fit to the audio signal. That means less search ways but it is necessary to exist into the system templates for every word, which overload utilization.

b) finding phonemes into the dictionary for the speaker language. The advantage is a reduced space for keeping information but the disadvantage is increased by the power calculation.

4) *Dependence of the speaker* is the main element to design and implement speech recognition system. The system can be:

a) independent by speaker and in this case there is a great resources consumption to convert all into dialect human speech.

b) dependent by speaker, using minimum resources but they require to educate systems for a few hours and so accuracy is ninety percentage. The users with mobility disabilities prefer these systems because they are easily to use.

c) adjusted to speaker, meaning to educate system for the same speaker.

The speech recognition system must achieve an equilibrium among four elements and to assure independence by characteristics of voice users.

2. Text-to- speech (TTS)
Starting from a written text it suppose different constrains regarding vocabulary (theoretically unlimited) but pronounced sentences must be as naturally respecting an usual intonation. Because sentences can't be memorized, it's necessary to choose a limited en-

semble of linguistic units which by the process of concatenate allows the vocal synthesis from written text.

A text-to-speech system contains two units:

1. a unit regarding editing text, which makes all the operations meaning content analyze to convert text in audio codes.

2. a synthesizer, which contents hardware support for synthesis proceeding.

To create and use the rules for convert text into vocal signal are analyzed the following elements:

a) the phonemes (sound's parts which compose the words) used to produce computerized voice ;

b) the quality of vocal signal which depends on the rules of finding and converting the text to audio signal; intonation, emotions and some problems in pronunciation overload producing vocal signal. To eliminate these impediments it can be used tags control or associate transcription text version to transcription phonetics word.

c) text-to-speech through concatenate diphones ;

d) using phonemes pairs to produce each sound.

It's evaluated every word and then the system joins phonemes to pronounce the word. The system which concatenates diphones is individual language because phonemes are differentiated by speaker's language.
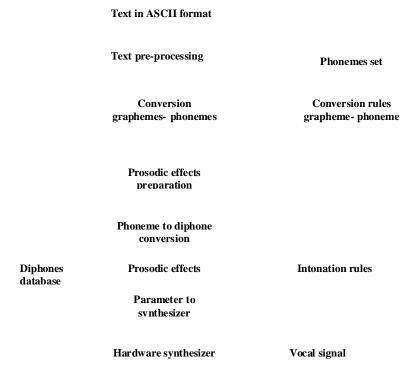
Text in ASCII format

Text pre-processing   **Phonemes set**

Conversion   Conversion rules
graphemes- phonemes   grapheme- phoneme

Prosodic effects
preparation

Phoneme to diphone
conversion

Diphones   Prosodic effects   Intonation rules
database

Parameter to
synthesizer

Hardware synthesizer   Vocal signal

Fig.2. Elements of TTS system for Romanian language

Speech synthesizers are the highest level of text-to-speech systems. They make the same thing as text-to-speech but there are the following parameters sets:

- source characteristics: basic frequency, volume, intensity
- acoustic characteristics of voice: the frequencies of high sonority, the band-pass width, the coefficients of prediction, the energy of the signals in of the filters.

MBROLA is an example for the speech synthesizer, which converts the text to sound using phonemes dictionary of Romanian language.

There are applied different accessibility technologies to make speech recognition and text-to-speech.
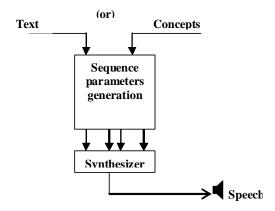
Fig.3. Difference between TTS system and speech synthesizer

3. SAPI Model (Speech Application Programming Interface)

Specific to Windows operating system allows speech recognition and conversion of the text in vocal signal. SAPI is used as interfaces COM (Component Object Model) and contents two distinctive levels:

- high SAPI, which allows the access to basic services of speech recognition by Voice Command Object and simple outputs text-to-speech through Voice Text object;

- low SAPI, which offers large access to Windows speech recognition services by SR Engine Enumerator Objects meaning SR Sharing and conversion of the text in vocal signal through TTS Enumerator and TTS Engine Objects.

Low level of SAPI services is useful for developed use speech recognition and TTS services.

4. Application for reading information on Internet (searching by word) using web services or speech synthesis from text files, Word and Excel

To accomplish such application the **tech-**
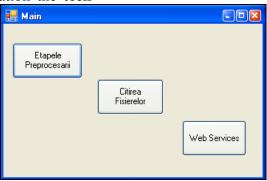**nologies** used are:

- Microsoft. NET Framework 2.0 package
- MySQL 4.1 (in which is made database)
- MyODBC 3.5.1 (for connection to MySQL)
- Programming languages to access database: C, C++, Java, PHP
- Audio programs for recording, processing and editing wave
- MBROLA synthesizer.

The **interface** with user contains:

• Levels of pre-proceeding text
• Reading information from files
• Accessing and reading information from web services.

Pre-proceeding text requires:

- connection to database by phonemes and waves
- use phonemes dictionary of Romanian language (from MBROLA synthesizer)
- pre-proceeding of the text which can contain letters and numbers;
- conversion of the text in .pho format, which can be read by MBROLA;
- conversion in wave format.

Reading information from Word or Excel files requires adding COM references from working with them. File reading is made by writing its name and extension into Text Box Audio.

In application's interface was added the option to print the text in Braille format and print preview before printing, because learning by two channels, hearing and touching, assures to memorize the information for a long period of time.

Conclusions

For creating software products, the developers must take in account compatibility between their products and accessibility standards.

The accessibility, for persons with disabilities, means the use of the instruments which have on its basis the speech synthesis.

Reproducing of a natural voice remains an open path for research and a challenge for software designers.

References

[1] Amundsen, M., *MAPI, SAPI and TAPI Developer's Guide*, SAMS, 1996;

[2] Pollard, C, Sag, I.A. - *Head-driven Phrase Structured Grammar*, CSLI, 1994;

[3] Reveiu, A., Dârdală, M., *Designing of accessible software for economic and social inclusion*, The Proceedings of the 6th Biennial International Symposium, SIMPEC 2006, Braşov;

[4] Smeureanu, I., Dârdală, M., Reveiu, A., *VISUAL C# .NET*, Editura CISON, Bucureşti, 2004;

[5] MBROLA: *http://tcts.fpms.ac.be/*

[6] Ferency, A., Zaiu, D., Ferencz, M. , Raţiu, T. *Sistem Text-To-Speech pentru Limba Română*; *http://www.racai.ro/*

[7]***, Accessibility in Microsoft Products *http://www.microsoft.com/enable/*