SYSTEMS OF CONTINUOUS SPEECH RECOGNITION FOR ROMANIAN LANGUAGE

Cătălin Chivu

"Transilvania" University of Brasov, Department of Economic Engineering and Manufacturing Systems

***Abstract**: The speech command has, lately, increased its applicability fields. The continuous speech command systems for Romanian language are, unfortunately, still very rare systems. This is caused by the difficulty of design a continuous speech recognition engine specific to Romanian language. The present paper proposes an adaptation, for Romanian language, of a continuous speech recognition engine, designed for American English and an application of this new speech recognition engine for a hand prosthesis.*

***Keywords**: continuous speech recognition, hand prosthesis, Romanian language, Grammar.*

1. INTRODUCTION

Speech recognition for Romanian language can be done by using a speech recognition system specially designs for this task. Designing such a system implies to create a specific corpus that contains the training data set and also the testing data sets, to create (from the beginning) a recognition systems and to train and to integrate this system in some specific applications. The system designed is one for continuous speech recognition and that because the goal is to be a high applicable system. The recognition engine, by itself, represents a program entity that is sold by the software companies, companies that invest fabulous sum of money, for many years, in specialists' teams that should design, test and develop such systems. The product of such a company is a package of software in their executable form (source cod never leaves the company). Thus, the idea that one person (or one specialist) could develop by himself such a system is a utopia.

The alternative solution is to consider an existing recognition system, which has satisfactory performances, even if it was already particularized for other language, and to adapt it and force it to recognize Romanian language. To adapt the system it should be replaced the grammar of the recognition system (the grammar is designed for applications in English) with other grammar, special designed for the desired application in Romanian language. To rewrite the grammar of the native language of the system for other language, different of the native one, implies, among redesigning the desired expression, according to the rules assess by the system, the phonetic transcription of each Romanian word, which is part of the new grammar, based on the phonemic collection of the native language of

the recognition system (in this situation the American English language).

2. TECHNIQUES TO IMPLEMENT THE PHONEMIC REPRESENTATION FOR ROMANIAN LANGUAGE

Although the English language is not a phonemic one, most of the words having a pronunciation and writing totally different, the phonemic writing rules of English, directly link the pronunciation and the writing of the words.

Strictly analyzing the phonemic collection of the SAPI5 speech recognition package, it can be observed certain similarity with Romanian phonemic collection.

The *aa* phoneme form English word *father* is the same from Romanian word *ladă*. This remark can be proved by analyzing the two spectrograms (first of English word *father* and second of Romanian word *ladă*) represented in Fig. 1 and 2.

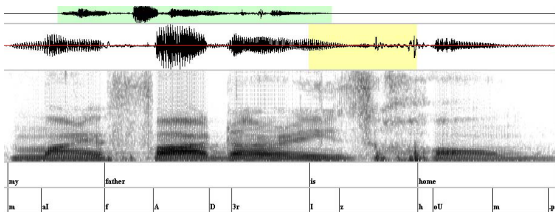

Fig. 1. The phonemic labeling for the expression "my father is at home"

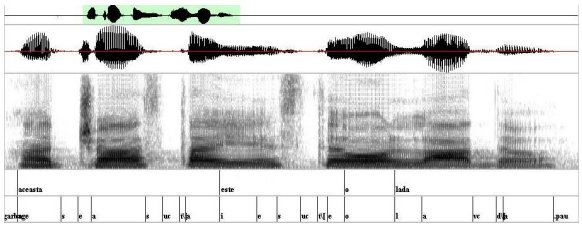

Fig. 2. The phonemic labeling of expression "aceasta este o ladă"

The *ax* phoneme from English word *ago* is the same with that from Romanian word *hău*.

The *b* phoneme from English word *big* is the same with that from Romanian word *stabil*.

The *ch* phoneme from English word *chin* is the same with that from Romanian word *cină*.

The *d* phoneme from English word *dig* is the same with that from Romanian word *dig*.

The *eh* phoneme from English word *pet* is the same with that from Romanian word *pet*.

The *er* phoneme from English word *fur* is almost the same with that from Romanian word *măr*.

The *f* phoneme from English word *fork* is the same with that from Romanian word *foc*.

The *g* phoneme from English word *gut* is the same with that from Romanian word *gaz*.

The *h* phoneme from English word *help* is the same with that from Romanian word românesc *han*.

The *ih* phoneme from English word *fill* is the same with that from Romanian word *stil*.

The *jh* phoneme from English word *joy* is the same with that from Romanian word *imagine*.

The *k* phoneme from English word *cut* is the same with that from Romanian word *cal*.

The *l* phoneme from English word *lid* is the same with that from Romanian word *lin*.

The *m* phoneme from English word *mat* is the same with that from Romanian word *mal*.

The *n* phoneme from English word *no* is the same with that from Romanian word *nărav*.

The *ow* phoneme from English word *go* is almost the same with that from Romanian word *nod*.

The *p* phoneme from English word *put* is the same with that from Romanian word *puţ*.

The *r* phoneme from English word *red* is almost the same with that from Romanian word *rest*.

The *s* phoneme from English word *sit* is the same with that from Romanian word *dosit*.

The *sh* phoneme from English word *she* is the same with that from Romanian word *şiret*.

The *t* phoneme from English word *talk* is the same with that from Romanian word *talc*.

The *uh* phoneme from English word *book* is the same with that from Romanian word *suc*.

The *v* phoneme from English word *vat* is the same with that from Romanian word *vară*.

The *z* phoneme from English word *zap* is the same with that from Romanian word *zar*.

The *zh* phoneme from English word *pleasure* is the same with that from Romanian word *jar*.

However, it can be observed that phoneme *î* or *â* (from Romanian word *cât*) is missing. That is why, this phoneme will be substitute by *ax*

phoneme, presented above, which is the most closed, in pronunciation. The other Romanian phonemes that have no correspondent could be described by using combinations of two or more English phonemes.

The *ch eh* phonemes are equivalent to the phoneme corresponding to *ce* group of letters from the word *cenuşă*. The *k eh* phonemes are equivalent to the phoneme corresponding to *che* group of letters from the word *cheie*. The *k ih* phonemes are equivalent to the phoneme corresponding to *chi* group of letters from the word *schit*. The *d jh ae* phonemes are equivalent to the phoneme corresponding to *ge* group of letters from the word *gel*. The *g y eh* phonemes are equivalent to the phoneme corresponding to *ghe* group of letters from the word *ghem*. The *d jh ih* phonemes are equivalent to the phoneme corresponding to *gi* group of letters from the word *gin*. The *g ih* phonemes are equivalent to the phoneme corresponding to *ghi* group of letters from the word *ghid*. The *t z* phonemes are equivalent to the phoneme corresponding to *ţ* letter from the word *ţap*.

Such a phonemic equivalence works perfectly because the recognition system decomposes the vocal signal in frames, gives probabilities to each phoneme for each frame and then, the Viterbi search selects the words that have the highest total probability. A compose phoneme of Romanian language will be automated decompose in phonemes specific to recognition system, when the signal is decomposed in frames. If the phoneme from Romanian language is very precisely described, using the recognition system phonemes, it allows it to get a high probabilistic score and, in consequence, the word can be correctly recognized.

3. VOCAL COMMAND APPLIED TO CONTROL SYSTEM FOR A HAND PROSTHESIS

The present chapter of the paper presents the structure of a complex system, which uses the speech recognition in Romanian language for control hand prosthesis. The grammar of the speech recognition system was written thus it allows a variety of commands. The commands were classified on two levels of complexity.

The first level of complexity corresponds to cosmetic commands that are dedicated to the control of the functional parameters of the prosthesis, as gripping force, angular speed of joints (low level commands). Some of these commands imply the numeral, which also represents a complex grammatical section.

The second level of complexity corresponds to the simple direct commands (high level commands) that are dedicated to complex movements.

A very important goal is that to create the high level commands as simpler as possible (for example: the command *"prinde creion" – catch the pencil* – that should determine the movement of the prosthesis in a predefined position, its rotation, the catch of the pencil and orientation of the pencil in the writing position). The simplicity is very important for the ergonomics of the prosthesis. As simpler the commands are, as efficient the speech recognition system becomes and the prosthesis is easier to be used.

3.1. Grammar section of the prosthesis

The grammar special created for the vocal command of prosthesis has a structural form, which gives a very high flexibility from composition of commands point of view. Thus, as example, it is presented a section of this grammar.

```
<RULE NAME="comenzi_rapide" TOPLEVEL="ACTIVE">
<L PROPNAME="comenzi_rapide" PROPID="PID_comenzi_rapide">
<P>
<RULEREF NAME="rscurte" PROPNAME ="pscurte" PROPID="PID_pscurte"/>
</P>
<P>
<P>
<RULEREF NAME="rcompuse" PROPNAME ="pcompuse" PROPID="PID_pcompuse"/>
</P>
<O>
<RULEREF NAME="robiecte" PROPNAME ="pobiecte" PROPID="PID_pobiecte"/>
</O>
</P>
</L>
</RULE>
```

This section defines a rule called *"comenzi rapide"* (*fast commands*) that has a list of phrases also called *"comenzi rapide"* (*fast commands*). If the rule has the parameter TOPLEVEL="ACTIVE" then the speech recognition system verifies each vocal expression if there is created based on the

topology of this rule. After checking, the expression can be recognized as being or not being created based on the topology of this rule. The list contains two phrases. The first phrase is, in fact, a reference to other rule, called "*rscurte*" that describes the short commands that can be done by the prosthesis. The second phrase is composed from a compulsory one (in fact a reference to another rule called "*rcompuse*" that describes the compose commands) and an optional section (which is also a reference to the rule called "*robiecte*" that describes the list of objects of the compose commands).

A command to the system may have three solutions: may be recognized as a short command, may be recognized a compose command with or without target object or may be rejected by the system as not being part of the list of accepted command.

In the following, it is presented the grammatical rule that describes the list of short commands accepted by the system.

```
<RULE NAME="rscurte">
<L PROPNAME="scurte" PROPID ="PID_scurte">
<P VAL="1">/inchide aplicatia/inchide aplicatia/ax n k ih d ae aa 1 p l iy k aa 1 t z iy aa 1;</P>
<P VAL="2">/stop/stop/s t ow p ;</P>
<P VAL="3">/reset/reset/r ae s ae t ;</P>
<P VAL="3">/initializare/initializare/iy n iy t z iy aa 1 l iy z aa 1 r ae ;</P>
<P VAL="4">/elibereaza/elibereaza/ae l iy b ae r ae aa 1 z ax 1 ;</P>
<P VAL="5">/indica/indica/iy n d iy k ax 1 ;</P>
<P VAL="6">/flexie/flexie/f l ae k s iy ae ;</P>
<P VAL="7">/extensie/extensie/ae k s t ae n s iy ae ;</P>
<P VAL="8">/opreste/opreste/ow p r ae sh t ae ;</P>
<P VAL="9">/continua/continua/k ow n t iy n uw 1 ax 1 ;</P>
<P VAL="10">/noroc/noroc/n ow r ow k ;</P>
<P VAL="11">/cere/cere/ch ae r ae ;</P>
<P VAL="12">/ofera/ofera/ow f ae r ax 1 ;</P>
<P VAL="13">/perfect/perfect/p ae r f ae k t ;</P>
<P VAL="14">/apasa buton/apasa buton/aa 1 p aa 1 s ax 1 b uw 1 t ow n ;</P>
<P VAL="15">/varsa/varsa/v aa 1 r s ax 1 ;</P>
</L>
</RULE>
```

Rule called "*rscurte*" has a list of phrases (the properly commands). One of these phrases may correspond to a part of vocal expression. Each of these phrases has a numeric parameter, called VAL, which is returned to application by the recognition engine when the expression corresponds to a part of the analyzed vocal expression. The syntax of the rules includes the name of the commands, the returned text value and the desired pronunciation (phonetic transcription of the vocal segment that may correspond to the respective phrase).

The grammar specific to prosthesis contains also the rules corresponding to the compose commands, the list of the objects that can be manipulated and the list of numeral. The numeral is also a very complex grammar structure that allows recognition of numbers between 0 and 999,999, spelt according to the pronunciation rules of numeral in Romanian language or according to usual abbreviated pronunciation.

To obtain the phonetic transcription of the command in Romanian language were used the phonetic transcription rules created for the vocal synthesis system.

3.2. Application of the vocal command of prosthesis

The module of results evaluation implies the implementation, in some concrete application, of the recognition engine, the grammar specific to the command of prosthesis and the prosthesis itself, together with its controllers and the predefined specific set of movements. The evaluation is done by testing the application that does the vocal commands recognition, their decomposition, their treatment and the simulation of predefined movements, according to the command. Simulation application of vocal commands of prosthesis, developed to evaluate the performances of such a system, has many functional blocks as it can be seen in Fig. 3. The window of the vocal command application of prosthesis is presented in Fig. 4.

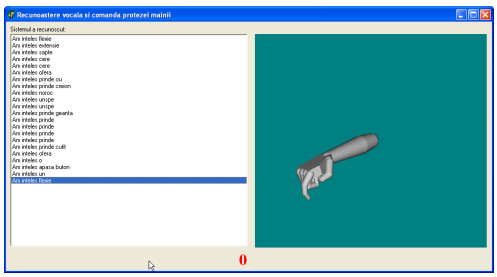

Fig. 3. The window of application of prosthesis simulation

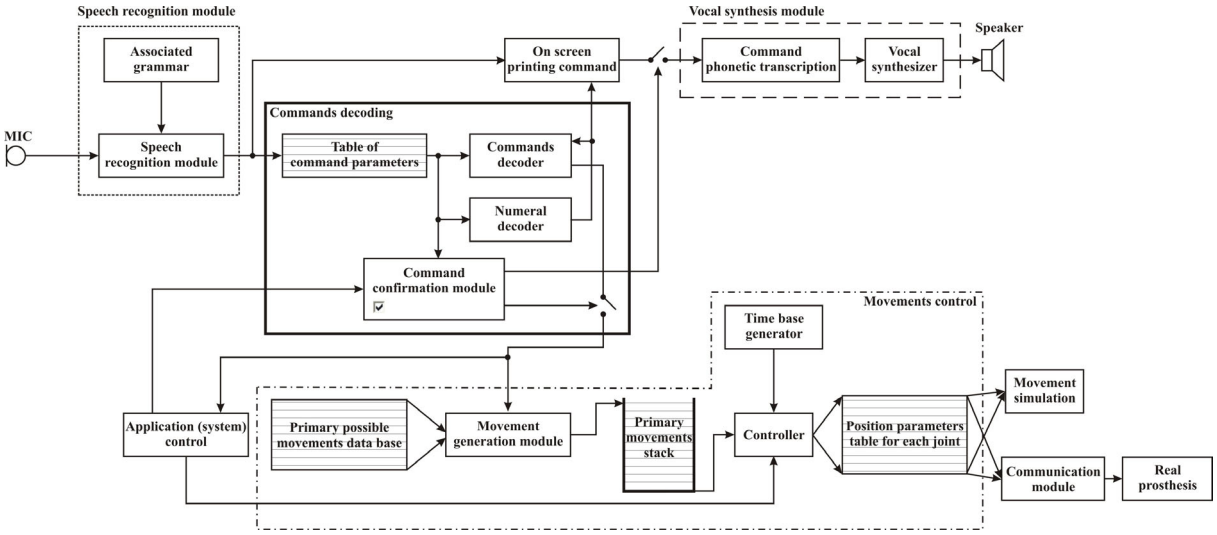

Fig. 4. The diagram of the vocal command system of the prosthesis

The user pronounces a command (an expression), in Romanian, that should be included in the flexibility limits given by the system grammar. The vocal signal is processed by the speech recognition system that, using the grammar, was particularized for the Romanian language speech recognition and for the prosthesis command application. The speech recognition module transmits to application the identification values of the recognized key words, of the recognized phrases and their text. The application will show on screen the text recognized by the module of speech recognition and puts the transmitted values into a command parameters matrix.

The commands decoder analyses the values from the command parameters matrix and, based on them, transmits to the module that generates the movement or to the prosthesis control module, the operation that is wanted to be executed. Further, the movement module receives the primary movements that create the desired complex movement, from a matrix of possible predefined primary movements and send them to a stack of primary commands selected by an appropriate order.

The prosthesis controller, coordinated by a base time generator, extracts, one by one, the primary commands from the stack and transforms them in movement information (position, speed, force), which is transmitted, to by executed, to a registers matrix that contains the movement and position parameters for each joint of prosthesis.

Simulation module receives the parameters from the movement parameters matrix and positions the visual elements of prosthesis according to them.

3.3. Advantages and disadvantages of the vocal command of prosthesis

The main advantages of the vocal command system of the hand prosthesis are:

- high ergonomics;

- the specific grammar of the system can be strongly personalized for the patient, after criteria of diagnostic, activity, dialect and vocabulary. Further, the prosthesis can be adapted to the voice tonality of the same patient during life time;

- the system can be, practically, improved with a number, practically unlimited, of commands and variants of them together with their specific movement parameters;

- the vocal command system is practically applicable for any type of structure of prosthesis. Furthermore can be used to upper limb prosthesis;

- possibility of activating the module of command confirmation allows the use of it even in noisy environment;

- compare to the myoelectric command of prosthesis, a very modern command method, the vocal one does not need the attached

electrodes and the corresponding stimulation system, positioned on the healthy part of the limb. Furthermore, the myoelectric one, after a period of time, determines uncontrolled muscular movement and thus, the inaccurate function of prosthesis and a discomfort, comparative to the vocal command that is qualitatively constant from movements' point of view.

- the prosthesis command do not imply odd movements of the body;

- once the system is trained for one patient, the vocal command system, being unnecessary to adapt to many speakers, it can attained a recognition rate very close to 100%.

The main disadvantages of the vocal command system of hand prosthesis are:

- the environment should be noiseless, thus the system will correctly recognize the desired commands;

- the vocal command system should identify the speaker by his voice parameters;

- the vocal command implies a delay proportional with the desired command complexity (the system can not decide until the speaker finishes the expression of the desired command);

- in some circumstances, when the system probably confuses the commands (because of the noise or voice parameters changes), patient should be able to chose to confirm the commands (but this aspect increase the reaction time);

patient should always wears microphone headset.

4. CONCLUSION

By adapting the grammar and the phonemic rules of the speech recognition engine for American English, to Romanian language, was obtained a new speech recognition system. The validity of this adaptation was verified by the application presented in the present paper (a hand prosthesis speech command).

REFERENCES

[1] Abdelatty Ali, A.M., Spiegel, &co – „An acoustic-phonetic feature-based system for the automatic recognition of fricative consonants". In Proc. ICASSP-98, pp. 961-964, 1998.

[2] Barto, A., Anandan, P. – „Pattern Recognizing Stochastic Learning Automata". In Proc. IEEE 15, pp.360-375, 1985.

[3] Brown, P. – „The Acoustic-Modelling Problem In Automatic Speech Recognition". PhD Thesis, Carnegie Mellon University, USA, 1987.

[4] Deng, L., Sun, D. – „Speech recognition using atomic speech units constructed from overlapping articulatory features". Proceedings Eurospeech-93, pag. 1635-1638, Berlin, Germany, 1993.

[5] „Dicţionarul ortografic, ortoepic şi morfologic al limbii române". Ed. Academiei Române, Bucureşti, 1982.

[6] Dumitrache, I., Dragoicea, M. – „Intelligent Techniques for Cognitive Mobile Robots". Revista „Control Engineering and Applied Informatics", no. 2, vol. 5, pag. 3-8, ISSN 1454-8658, 2004.

[7] Hieronymus, J. – „ASCII Phonetic Symbols for the World's Languages: Worldbet". AT&T Bell Laboratories, Murray Hill, USA.

[8] Kent, R. D., Read, C. – „The acoustic analysis of speech". San Diego: Singular Publishing Group, pag. 105-121; 129-130, 1992.

[9] King, S., Stephenson, T., Isard, S., Taylor, P., Strachan, A. – „Speech recognition via phonetically featured syllables". In Proc. ICSLP-98, Sydney, Australia, 1998.

[10] Ohde, R. N., Sharf, D. J. – „Phonetic analysis of normal and abnormal speech". New York: Merrill, pag. 55-93, 1992.

[11] Rabiner, L.R., Schafer, R.W. – „Digital Proccessing of Speech Signals". Prentice Hall Inc., New Jersey, 1983.