System of Speech Access in the Azerbaijani Language to Objects of Electronic Map

Alexander Abramenkov[1], Nina Petuhova[2], Mais Farkhadov[3]
Institute of Control Sciences of RAS, Moscow, Russia
[1]*aabramenkov@asmon.ru*, [2]*nvpet@ipu.ru*, [3]*mais@ipu.ru*

***Abstract* — This report describes the implementation of a speech access system for objects of electronic map. This system is works with the use of the computer-based recognition of words and phrases in the Azerbaijani language. The Recognition process is based on the recognition of another language which is phonetically close to Azerbaijani. For the purpose of creating a speech interface, a scenario of a dialog with the user has been developed, a dictionary of phonetic transcriptions has been compiled and the grammar has been defined. The integration with the controls of the electronic map of the city was has been adapted to become a new part of the existing comprehensive security system.**

Keywords— speech recognition; voice user interface; Hdden Markov Models; phonetic transcription; electronic map

I. INTRODUCTION

The development of a basic recognition system for any language is very time-consuming and costly. The process of designing a recognizer which is based on the Hidden Markov Models (HMM) schematically involves the following steps [1-8]:

- Carrying out a language features analysis and a phoneme classification. Selecting units of speech recognition and selecting the type of a model for a basic speech unit.
- Creating a phonetic dictionary. This dictionary must include such words and their combinations that would reflect most of the phonemes and their combinations presented in the language.
- Creating a speech database (a speech corpus) on a basis of a phonetic dictionary for the purposes of training and testing. In order to create a speech database, words and phrases from a phonetic dictionary must be recorded by different speakers, in different environments and with different microphones or telephones. Audio files are then marked and provided with a phonetic transcription. Part of the speech corpus (usually a large part) is used for training models, and the other part is used for testing.
- Carrying out a spectral analysis and the parameterization of speech signals from a training set of records of the speech database in order to obtain standard feature vectors and to train acoustic models which describe the elements of speech.
- Making the HMM's for all elements of speech and all words from the dictionary. Initial parameters of HMM's are then selected. After that, models must be trained. In the process of training HMM's parameters are sequentially reevaluated and the models with new parameters are tested on a test sample until the desired result is attained.
- The laboriousness of creating recognizers leads to the fact that for now recognizers have been created only for the world's most popular languages and dialects. Recognizer for the Azerbaijani language is not among them. Due to the current situation, we have decided to utilize the idea of using the phonetic and linguistic model of recognition which is designed for another language (phonetically close to Azerbaijani).

II. DESCRIPTION OF DEVELOPMENT

The base recognition pack for the Turkish language has been chosen to create the speech access system for objects of the electronic map of Baku. This choice is explained by the fact that both languages (Turkish and Azerbaijani) belong to the Turkic group and, therefore, have many features in common, such as most of the sounds and the pronunciation. Besides, their alphabets differ only by one letter.

Experiments have been performed using the Speech Recognizer Nuance Recognizer 9 [9] for the Turkish language. We have selected 100 objects on the electronic map of Baku and defined the appropriate grammar. There is no Azerbaijani letter '*Ə'* in the Turkish alphabet, so firstly we replaced it with letter 'e'. After that, on the software basis, we generated an auto-transcription dictionary according to the internal rules of the phonetic transcription of the Turkish language. The further analysis of that dictionary has revealed that the majority of all phonetic descriptions of words adequately have reflected their sound in the Azerbaijani language. However, the transcriptions for the 16 words out of 100 have had to be changed. For example, the Turkish word '*heyder*' pronounced in Turkish like '*hEdEr'*, i.e. the letter '*y'* is omitted while pronounced, but the same word '*heyder'* in the Azerbaijani language sounds like '*hEjdEr'*, i.e. the letter '*y'* must be present in the phonetic transcription.

Table 1 shows a list of words for which corrections have been made.

After making changes to the phonetic transcription the share of correctly recognized words have amounted to 0.94. This result has confirmed the feasibility and usefulness of the Turkish language pack for the Azerbaijan speech recognition.

TABLE I. LIST OF CORRECTED TRANSCRIPTIONS FOR STREETS OF BAKU

The word in the dictionary	**Auto-transcrip-tion**	**Corrected transcrip-tion**	**letter**	**Transcribe**	
				false	***correct***
heyder	hEdEr	hEjdEr	y		j
azadlıq	azadlIk	azadl1g	ı	I	1
mustafa	mUsdafa	mUstafa	t	d	t
babek	babEk	ba:bEk	a	a	a:
tbilisi	tbIlIsI	tIbIli:sI	i	I	i:
neftçiler	nEftdZIlEr	nEftSIlEr	ç	dZ	S
bağ	ba	bag	ğ		g
geray	gEra	gEraj	y		j
vurğun	vUrUn	vUrgUn	ğ		g
semed	sEmEd	samEt sEmEt	d	d	t
reşid	rESId	rESIt	d	d	t
behbudov	bEhbUdQv	pEhbUtQv	d	d	t
seferli	sEfErlI	safErlI	e	E	a
eliağa	ElIaa	ElIjaga	ğ		g
dairesi	darEsI	da:jrEsI	a i	a	a: j
ukrayna	Ukrajna	Ukra:jna	a	a	a:

User access objects of the electronic map with voice (by the use of a microphone or a headset with a wired or wireless connection). After a short dialogue, the address of the object in a textual form is generated. Our system supports the ability to clarify details of the location of the object. The result of the dialogue is passed on then to a dynamic management module that controls this object of the map.

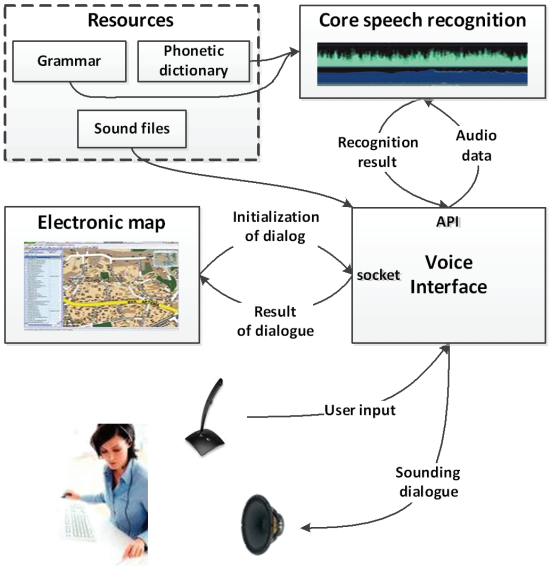

Figure 1. Architecture of the speech access system for objects of electronic map

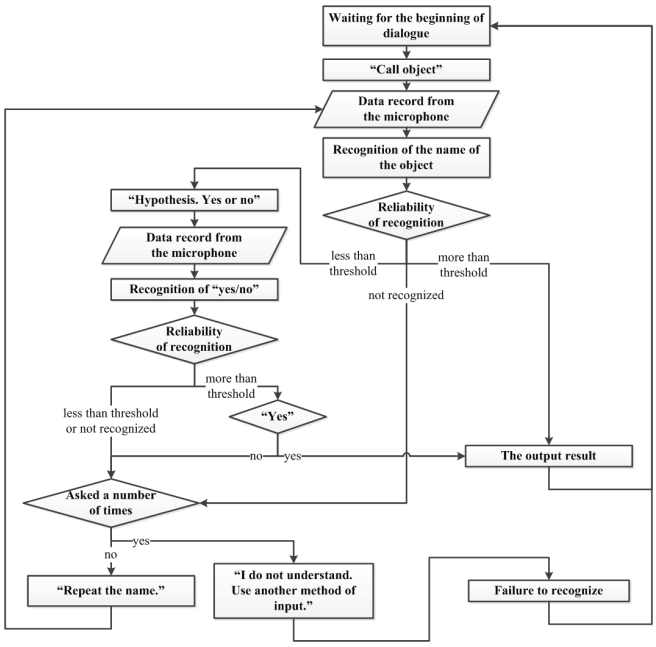

Figure 2. Flowchart for the speech access systems for objects of the electronic map

Fig. 2 shows a flowchart for the system. The development has been applied to the safety management system in Baku.

CONCLUSION

The speech interface for the electronic map of the city facilitates an operator's work and improves the efficiency of online monitoring which helps ensure the security of the city by making it possible to respond to emergencies rapidly. The speech interface for the security system provides an additional access channel to objects of the map and favors the improvement of operator's working conditions.

REFERENCES

[1] L.E. Baum and T. Petrie Statistical inference for probabilistic functions of finite Markov chain, Ann. Math. Stat., 37: 1554-1563, 1966.

[2] L.E. Baum, T. Petrie, G. Soules and N. Weiss A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, Ann. Math. Stat.., 41(1): 164-171. 1970.

[3] F. Jelinek, L.R. Bahl and R.L. Mercer Design of a linguistic statistical decoder for the recognition of continuous speech, IEEE Trans. Information Theory, IT-21: 250-256. 1975.

[4] F. Jelinek Continuous speech recognition by statistical methods, Proc. IEEE, 64: 532-536.

[5] L.R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proc. IEEE, 77 (2): 257-286, February, 1989.

[6] L.R. Rabiner, B.H. Juang Fundamentals of speech recognition, A Pearson Education Company Upper Saddle River, NJ. 1993 by AT&T. 507 p.

[7] Q.Y. Hong, S Kwong A training method for hidden Markov model with maximum model distance and algorithm, in Proc. of the Neural Networks and Signal, Prosecing, 465-468, Vol. 1. April 2004.

[8] M. Gales and S. Yuang The Application of Hidden Markov Models in Speech Recognition. Foundation and Trends in Signal Perosessing, Vol. 1, № 3 (2007), 195-304, 2008.

[9] www.nuance.com