Spoken Dialogue System for Mobile Parking

E.Meister[1], *P.Tatter*[1], *J.Lasn*[1], *R.Vahisalu*[2], *B.Lobanov*[3], *T.Levkovskaya*[3], *V.Kisialiou*[3]

[1]Institute of Cybernetics
Tallinn Technical University
Estonia, einar@ioc.ee

[2]Estonian Mobile Company
raul@emt.ee

[3]Institute of Engineering Cybernetics
Minsk, Belarus
lobanov@newman.bas-net.by

Abstract

Fast progress in mobile communication technology provoke new ideas and initiate research and development of innovative services and applications. This paper presents a prototype of a spoken dialogue system developed for parking system over mobile phone. The key issues of the dialogue system are a reliable word recognition and robustness against different channel distortions and speaker-specific variations. The voice input of the system is limited to spoken digits and names of Estonian letters. For voice output recorded prompts have been used. In the paper the word recognition algorithm, dialogue structure and test results are introduced.

1. Introduction

Car parking in city areas is nowadays a big problem, as a rule it is a paid service in most cities. There are several payment options – scratch tickets, parking machines, barrier systems, etc. – implemented in different parking areas. A modern trend is a chip card based parking systems which provide custom-friendly service with high security and electronically handled payments. At the same time these systems are technologically very complex and need big initial investment for additional infrastructure on streets and for the users. A novel approach is to take advantage of the mobile communication networks. The mobile parking system implemented in Tallinn, Estonia by the Estonian Mobile Company EMT has been launched in spring 2000. The service is available for all EMT clients, the service users should put a special sticker behind the front windshield. Parking account will be opened and loaded by calling a special service number 1902 (12 EEK will be loaded) or 1912 (75 EEK will be loaded). In order to start parking a customer should send a SMS messages including the license plate number and the zone code, eg. "123ABC central". The system returns a SMS message with starting time and balance. Ten minutes before running out of money a feedback SMS with a warning will be sent by the system. The customer should either leave the parking place in 10 minutes or to continue parking have a call to 1902 or 1912. When leaving the parking area a call to the service number 1903 stops the customer's counter and system returns a SMS message with start and end time, paid amount and current balance. Actual payment will be carried out with the next monthly bill.

The service is available via WAP interface, as well. Since launching the number of customers has been grown fast and reaches currently up to 10000 active users.

Payment inspection of parked cars is carried out by human inspectors who may connect the database through GSM voice session, SMS or WAP session, or using local radio session. For the further development of the service the use of a spoken dialogue system has been investigated. A dialogue system could be implemented in two ways:

- in the inspection process to assist the inspectors,
- in the parking process to assist the customers.

In both cases the vocabulary will be the same including up to 50 words (spoken numbers and letter codes, yes, no), but in the first case the number of speakers will be small (30-50) against a large number of speakers (10000-50000) in the second case.

Despite of a small vocabulary task there are demanding requirements for speech recognition: it should exhibit high performance and robustness against different kind of variations - different pronunciation of words by different speakers, variability of the speaker's voice, characteristics of the microphone and telephone lines, channel and background noises, etc.

A speech recognition system which should meet the above mentioned requirements has been developed by the researchers from the Institute of Engineering Cybernetics, Minsk, Belarus. The recognizer has been integrated into a spoken dialogue system. The voice input from the mobile network includes isolated spoken digits and names of Estonian letters. For the output recorded prompts have been used. The dialogue structure is based on a realistic model close to real parking process. For training and evaluation of the recognition system an application-specific speech database has been recorded, preliminary field tests have been carried out.

In the paper the speech recognition algorithm is introduced, the structure of the dialogue system, a dialogue model and the test results are presented, and approaches for further development are discussed.

2. Speech recognition algorithm

The basic speech recognition technique used in the prototype is a Continuous Dynamic Time Warping (CDTW) [1]. The main advantage of the CDTW algorithm is, that it evaluates the presence of a word (or subword) in a continuous speech signal and in addition, it evaluates the time of word's beginning, end and duration. The algorithm is based on a formant analysis method [2], which estimates the following sub-sets of formant parameters:

- F1(t), F1'(t), F2(t), F2'(t), F3(t), F3'(t) – the formant frequencies and their first derivatives;
- A1(t), A1'(t), A2(t), A2'(t), A3(t), A3'(t) – the formant amplitudes and their first derivatives;
- V(t), V'(t), E(t), E'(t) – voicing degree and frame energy of a speech signal and their first derivatives.

The algorithm implements the 'multi-stream' approach in a two-step word recognition procedure [3]. The following set of decision rules have been implemented:

1. CDTW evaluation of integral similarity for q-th word pattern – Sq(1);
2. Correlation degree of the acoustic parameters of whole word to be recognized and q-th word pattern – Sq(2);
3. Degree of temporal matching of the whole word – Sq(3);
4. Correlation degree of the left sub-word and q-th left sub-word pattern – Sq(4);
5. Correlation degree of the right sub-word and q-th right sub-word pattern – Sq(5);
6. Distance matching of the left sub-word and q-th left sub-word pattern – Sq(6);
7. Distance matching of the right sub-word and q-th right sub-word pattern – Sq(7).

At the first step a limited number of word candidates is provided by the 'erudite' recognizer. The decisions are ranked in the order of decreasing Sq(1) values. The number of candidates Q is found according to the formula:

$$Sq(1) > 0.7*Smax(1).$$

At the second step these candidates are examined by nine 'competent' recognizers using the rules 2-7. The final decision is achieved as a result of 'voting' by the competent recognizers according to their competence degree (weighting coefficients).

3. Structure of the dialogue system

The structure of the dialogue system is presented in Figure 1. In addition to the ASR-unit it includes the training unit and the dialogue unit. A prototype version of the system has been developed on PC-platform under Windows 98. The system is interfaced with telephone lines via TAPI-compliant voice modem. Currently, the prototype version is not integrated with the working parking system and there is no access to actual customers database. Instead a database emulation module has been developed which enables to test the dialogue system in near-real environment.

The training unit provides tools for speech recording and fast training options for a new user.

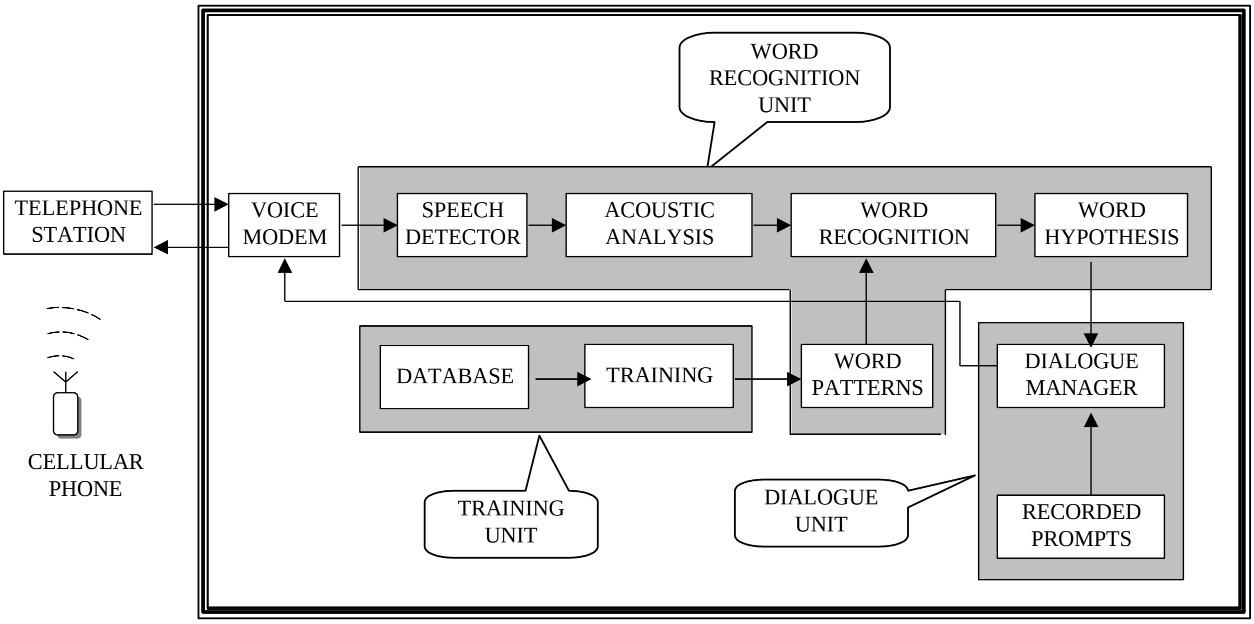

Figure 1. The structure of the dialogue system.

4. Database

An application oriented vocabulary (Table 1) was recorded for training and testing the speech recognizer. The database signals were collected over mobile phone from 10 male and 10 female speakers, 5 calls for each speaker from different locations (office, street). The signals were sampled at 8 kHz and stored in wav-format.

The database has been divided into two equal parts: half of the signals from every speaker have been used for training and the rest for evaluation.

Table 1. Vocabulary used by the recognition system

0 – null	A – Anna	N – Narva
1 – üks	B – Berta	O – Otto
2 – kaks	C – Cicero	P – Paul
3 – kolm	D – Dora	Q – Quebec
4 – neli	E – Emma	R – Richard
5 – viis	F – Friedrich	S – Salme
6 – kuus	G – Gustav	Z – Zürich
7 – seitse	H – Haapsalu	T – Teodor
8 – kaheksa	I – Iida	U – Urve
9 – üheksa	J – Jaan	V – Viktor
yes – jah	K – Karla	W – Washington
no – ei	L – Linda	X – Xenofon
	M – Marta	Y – Üpsilon

5. Dialogue structure

System: *Welcome to the mobile parking system*.

1) **System:** *After signal say the digits of your license plate number, please.*
System: <beep>
User: *FIVE*
System: <beep>
User: *SIX*
System: <beep>
User: *EIGHT*

System: *Your digits were: five six eight. Is that correct?*

User: *YES (or NO)*
In case of NO, system returns to 1. If user's answer is NO for three times, system goes to 3.

2) **System:** *After signal say the letter codes of your license plate number, please.*
System: <beep>
User: *ANNA*
System: <beep>
User: *KARLA*
System: <beep>
User: *SALME*

System: *Your letters are: anna karla salme. Is that correct?*

User: *YES (or NO)*
In case of NO, system returns to 2. If user's answer is NO for three times, system goes to 3.

System performs the database search and if the plate number is not found, the number will be stored into the database and the following message will be delivered:

System: *Parking counter for your car with the plate number five six eight anna karla salme is activated.*
System jumps to 3.

If the plate number has been found in the database, system delivers the message:

System: *Parking counter for your car is currently active, would you like to stop parking?*

User: *YES (or NO)*
In case of YES, system stops the counter, deletes the plate number from the database and informs user about elapsed parking time. Jump to 3.
In case of NO system goes to 3.

3) **System:** *Thank you, have a nice day.*

In case the speech detector is not able to detect signal in a specified time interval system gives the message:
System: *I didn't understand you, please repeat.*

6. Training and test results

Different training and recognition procedures have been tested. Both speaker-independent and speaker-dependent word patterns were created using the training set. The recognition test were carried out with the evaluation set and with real on-line signals from telephone channel (phone calls from different locations using different handsets and channels – fixed and mobile). The results are presented in the Table 2.

Table 2. Testing results (word recognition rates, %) of the word recognition system

	Speaker-independent patterns		Speaker-dependent patterns	
	Numbers	Letter codes	Numbers	Letter codes
Evaluation set of the database	100	98	100	100
On-line telephone calls	72	58	90	86

The test results show that the speaker-dependent word patterns outperformed the speaker-independent patterns. The word recognition rate in the case of on-line calls drastically dropped down comparing with the evaluation set. The most frequent errors are listed in Table 3.

Table 3. Most frequently confused spoken numbers and letter codes occurred in on-line tests

Actual digits and letter codes	Recognized digits and letter codes
1	5, 7, 9
3	2
4	5, 7
A	S
B	E, R, Z, X
C	R, X
D	J, X
E	X
F	S, Z
K	N, S
L	Z, X
M	H, N
N	J, S, T
O	G, H, Q, S
P	S, Z
Q	S, Z, X
T	J, S, Z, X
U	D, J, S
V	R
Y	R

7. Discussion

How to explain the drastic drop-down of recognition rates in realistic environment? Analysis showed that the main problems are the following:

- during database recordings and on-line tests different modem cards with different characteristics were used, i.e. the training and testing conditions were very different;
- acoustic characteristics of speech are influenced by different channel distortions and background noise which cause problems for speech detector;
- several speaker-independent word patterns occurred to be too generalized producing many confusions;
- before tests no adjustment between channel (modem card) and system input was carried out;
- several errors are produced by the speech detector which failed in determining of beginning and end of a word;
- there are several phonetically close word pairs, e.g. 'kaheksa-üheksa', 'Karla-Narva', 'Marta-Narva', 'Anna-Salme';
- in several cases the timing of subject’s reaction was not proper.

Test results with speaker-dependent patterns are much promising but should be improved, too. With the current recognition rate total failure of dialogues was 8%, rate of successful dialogues is given in the Table 4.

Table 4. Rate of successful dialogues

Number of repetitions	Success rate, %
No repetition	58
One repetition	25
Two repetitions	9

8. Conclusions

In the paper a prototype version of the dialogue system for mobile parking was introduced and the preliminary test results were presented. The prototype is in early development phase and several methods for further improvement are under investigation. Current recognition rates could be certainly improved by implementing of speaker and channel adaptation methods and by adequate modeling of realistic application environment.

9. References

[1] Lobanov, B., Levkovskaya, T. (1997) Continuous Speech Recognizer for Aircraft Application. – Proceedings of SPECOM'97, Napoca, Romania.

[2] Lobanov, B., Levkovskaya, T., Kheidorov, I. (1999) Speaker and Channel-Normalized Set of Formant Parameters for Telephone Speech Recognition.- Proc. of Eurospeech’99, Budapest, Hungary, vol.1, pp. 331-334.

[3] Lobanov, B., Levkovskaya, T. (2000) Multi-Stream Words Recognition Based on a Large Set of Decision Rules and Acoustic Features. - Proceedings of International Workshop “Speech and Computer” - SPECOM'2000, St.Peterburg, Russia, pp. 75-78.