

MARTA ŻAKOWSKA*

VOICE CONTROLLED HOME E-BANKING

STEROWANE GŁOSEM BANKOWE KONTO
INTERNETOWE

Abstract

Nowadays the progress in modern technologies in area of computer science and wide understanding communication, both using computer and communication with the computer, has a direct influence on our real life, institution and individuals. Rapid growth of the Internet and the easy access to it, makes that lots of information is available on the web. On the other hand, the Internet is not only for special groups of people. Nowadays more and more disabled people, also blind people, are using computers and the Internet. The adaptive technology industry is mainly focused on computer access, through screen readers and others specialized devices which give the blind people access to the sighted world.

In the article there is presented internet's bank application controlled by means of voice and keyboard, which helps disabled people in a prosaic tasks.

Keywords: ASR, TTS, application for disabled people, voice controlled system

Streszczenie

Obecnie postęp nowoczesnych technologii w obszarze informatyki i szeroko rozumianej komunikacji, zarówno przy użyciu komputera jak i komunikacji z komputerem, ma bezpośredni wpływ na nasze życie, ludzi i instytucje. Szybki rozwój Internetu i łatwy do niego dostęp sprawia, że wszelkie informacje są dostępne w sieci. Z drugiej strony Internet nie jest przeznaczony dla wybranych osób. Coraz więcej niepełnosprawnych ludzi w tym niewidomych i niedowidzących używa komputera jak również korzysta z Internetu. Dlatego też rozwój technologiczny ukierunkowany jest na ułatwienie dostępu do komputera, poprzez wprowadzenie syntezy mowy i innych wyspecjalizowanych urządzeń, które dają niewidomym dostęp do „świata widzących”. W niniejszym artykule jest przedstawiona aplikacja e-banku sterowana za pomocą głosu i klawiatury, która ma pomóc niepełnosprawnym w korzystaniu ze strony internetowej banku oraz serwisu bankowości elektronicznej.

Słowa kluczowe: system ASR, system TTS, system sterowany za pomocą głosu, aplikacja dla osób niewidomych

* MSc. Marta Żakowska, Institute of Applied Informatics, Faculty of Mechanical Engineering, Cracow University of Technology.

1. Introduction

Today lots of daily activities such as education, shopping, mailing, banking transactions are done through the Internet. Checking up rate of exchange or logging in to account is really easy to do in a few minutes on the bank's website. The Internet, is a window of the world which users are from different ages and groups. The disabled people also use the Internet. Some websites are especially created for them.

Voice controlling of the computer and applications, using speech are very promising fields, especially for disabled people. For those who have difficulties in entering data by means of the input tools like keyboard, mouse, etc., speech recognition is an effective alternative to alter or to combine input methods.

There are four main factors related with speech recognition problem [1]:

Speaker – a huge challenge is creating techniques that can accurately and reliably recognize anyone's voice and any dialect of a given language.

Coarticulation – the spectral characteristics of a spoken word (or sounds within the word) vary depending on what words (or sounds) surround it.

Speaking rate and style – it is difficult to get stable patterns for sounds or words that can be used with all speakers and speaking rates,

Environmental conditions – speech can be difficult to recognize in home environments (background speech from radios or TV).

These factors are sources of speech signal variability which must be carefully considered when developing applications based on speech recognition technology.

In the article an idea of internet's bank application operated by means of keyboard and voice is presented. There is described construction of voice controlled system composed of subsystems and adjusting website for blind and sand-blind.

2. Analysis of available solution

The given problem of making user-friendly internet application for blind people has some solution, but the obtained results of available solutions are not always satisfactory.

It is possible to find bank the website adjusted to people with sight dysfunction. Thanks to the sonorisation of whole service, using internet browser with the human voice and disposable codes in Braille's, for log in to the system and authorize transactions, bank gives possibility for blind to use all services and products self-dependently. However, data presented in [2] show that they are not the best solutions. Group of people who test bank's website the most difficulties met with log in form to the bank systems.

Almost all applications depend on a user's keyboard strokes and mouse clicks with a display monitor for feedback. Many browsers enabled transition to mark of reference by keyboard short cut defined on the website. One of browsers requires pressing simultaneously "Shift+Esc" keys, then appropriate access key. It is not comfortable for user. The solution is to simplified access key or to realize bank operations using a voice.

2. Project of application

Analyzing the most modern solutions in the area of voice controlled application, in the paper an original home e-banking application is proposed, as is presented in Fig. 1. The first step is to adjust website or on-line transactions service to the blind people needs. The second step is to create voice controlled system, which helps users to communicate with application by means of voice.

Assumption data are:

- 1) The system consists of microphone through which is applied input in the form of speech signal.
- 2) The ASR which is part of this project have the ability to adapt to a speaker. A system is trained by having the speaker repeat standard and common phrases. The advantage of this process is, that training a recognizer usually improves its accuracy. ASR systems with training should be able to adapt when the speaker consistently repeats an utterance.

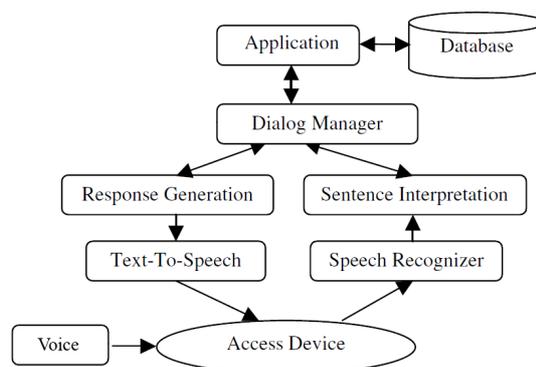


Fig. 1. Scheme of voice controlled e-banking application [4]

Rys. 1. Schemat sterowanej głosem internetowej aplikacji bankowej [4]

3.1. Website adjusting for blind people

Creating the legible bank website adjusted to the disabled people it is necessary to implement all possible facilities, what make easier for users.

The most important are:

1. To define key access enable to use the keyboard in the easiest way
 - Key 1 – The main page.
 - Key 2 – Log in to the transaction system.
 - Key 3 – Move to the content of the page.
 - Key 4 – Come back.
 - Key 9 – Contact with bank.
 - Key 0 – Statement of accessibility.
2. Text of the references are chosen in a special way – to be comprehensible even beyond the context of the sentence in which they are.

3. Links contain title attribute, which describes text of the references in details, unless link's words precisely describes target document by themselves (e.g. article heading).
4. Help in navigation.
Using semantic web indexes on the page. <h1> tags used for marking main heading, <h3> for subheading.
5. Layout of the page is described by Cascading Style Sheets (CSS).
6. Accordance with standards.
 - websites should fulfill the most important guidelines like W3C, which recommendations concern, inter alia creating description of graphic's elements or define titles of lines,
 - pages should undergo validation as XHTML 1.0 Strict.

3.2. Voice controlled system

A spoken language system has one of the following three subsystems:

Speech recognition system – convert speech into words.

Speech synthesis system – convert written input to spoken output by automatically generating synthetic speech.

Spoken language understanding system – maps words into actions and that plans system-initiated actions.

There is significant overlap in the fundamental technologies for these three areas.

3.2.1. Speech recognition system

The ability of a recognizer can be examined by measuring its accuracy – or how well it recognizes utterances. This includes not only correctly identifying an utterance but also indicates when the spoken utterance is not in its vocabulary. The acceptable accuracy of a system really depends on the application.

Creating good speech recognition system is the first step in communications process by means of voice. On this stage the noise reduction is very important.

In order to diminish the computational effort and to send more reliable data to the recognition system, voice signal is passed to the pre-emphasis filter. This filter allows boosting the high frequencies of the signal, leaving the lower frequencies untouched for better processing without altering the integrity of the data. The pre-emphasis filter is implemented following the equation [3]:

$$S(n) = X(n) - C.X(n-1) \quad (1)$$

where:

- C – pre-emphasis coefficient,
- $X(n)$ – input sample,
- $S(n)$ – filter output.

Speech recognition system can be divided on three stages, it is shown in Fig. 2

Signal processing:

The speech signal is processed in the signal processing module that extracts spectral feature vectors for the decoder.

Recognition stage - Recognize individual speech.

The next stage is recognition of phonemes, groups of phonemes and words. This stage can be achieved by HMM (hidden Markov modelling), expert systems and combinations of these techniques. HMM-based systems are currently the most commonly used and most successful approach.

Decoding stage:

The last stage providing a search through a huge space of potential "source" sentences and choosing the one which has the highest probability of generating this sentence. It is used models to express the probability of words.

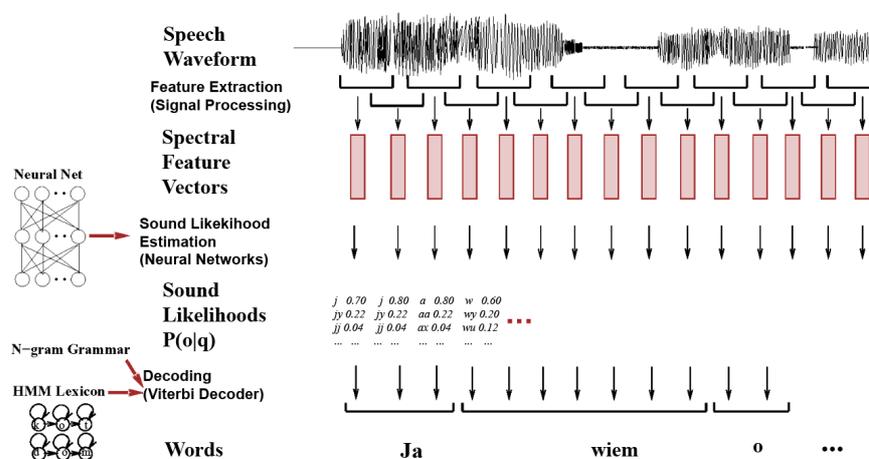


Fig. 2. Scheme of speech recognition

Rys. 2. Schemat architektury rozpoznawania mowy

Very important part in ASR is training a Speech Recognizer. It should be defined model to be trained and training data as below.

Models to be trained:

- language model: $P(w_i | w_{i-1} w_{i-2})$,
- observation likelihoods: $b_f(o_t)$,
- transition probabilities: a_{ij} ,
- pronunciation lexicon: HMM state graph structure.

Training data:

- collection of speech wavefiles + word-transcription,
- large text collection for language model training,
- smaller collection of phonetically labeled speech.

3.2.2. Speech synthesis system

Speech synthesis is in other words Text-to-Speech conversion. The task of a text-to-speech system can be viewed as speech recognition in reverse. Text is input by operator or by system OCR (Optical Character Recognition), purposive to optic sign recognition. In the Text-to-Speech conversion process it is possible to separate a few stages, which are successive elements of the speech synthesis system (Fig. 3).

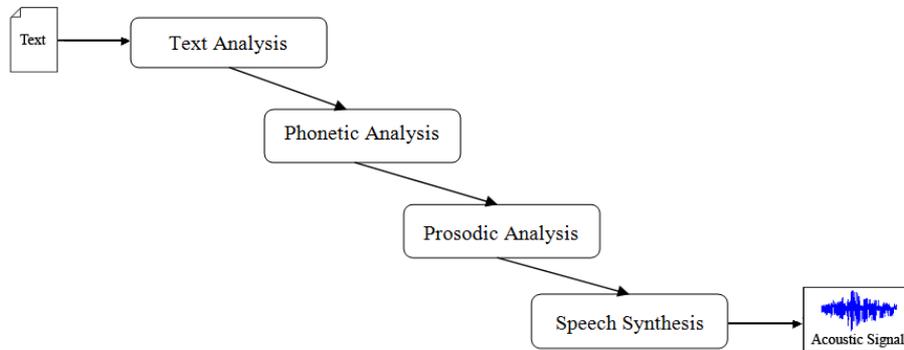


Fig. 3. Speech synthesis system structure

Rys. 3. Struktura systemu syntezy mowy

The text analysis module, which scheme is shown in Fig. 4, define type and structure of processed document. It makes also non-orthographic signs conversion, syntactic and lexical analysis.

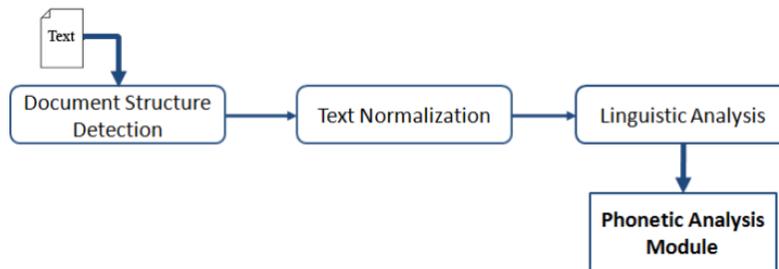


Fig. 4. Text analysis module

Rys. 4. Moduł analizy tekstu

Functionality of the module is as follows:

- Recognize punctuation signs.
- Recognize numbers – conversion number in figures on number in words.
- Recognize abbreviations of expressions, symbols.
- It is possible to add new abbreviations and symbols and edit already existed system.
- Mark all points in text, which gives information for prosodic analysis module – to insert tags in text.

Information processed in text analysis module are input for phonetic analysis module.

The aim of phonetic module is to convert words given as a orthographic code into phonetic code with additional information which define their pronunciation.

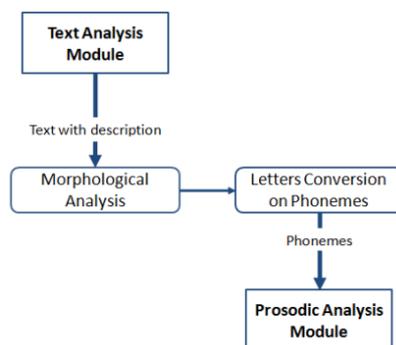


Fig. 5. Phonetic analysis module

Rys. 5. Moduł analizy fonetycznej

Phonetic analysis module has functionality as follows:

- Exactly define using phonetic transcription, transcription scheme is based on SAMPA.
- Recognize orthographic signs sequence consist on one phoneme’.
- Replace orthographic signs adequate phoneme symbols.

Data processed in this module are input data for prosodic analysis module.

Prosody study function accomplished by phonic features, which are not phoneme’s attributes, but attributes of syllables or syllables sequence. It includes accent, quantity and intonation. Performance of this module has very important influence on precision in simulating natural speech signal by synthesis speech module.

Prosodic model includes data for synthesis speech module.

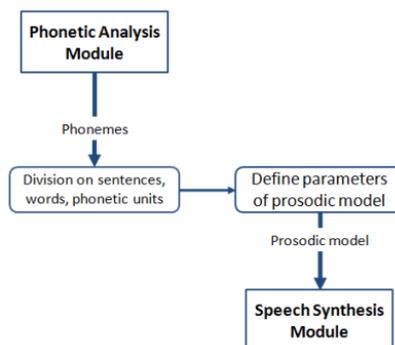


Fig. 6. Prosodic analysis module

Rys. 6. Moduł analizy prozodycznej

The last stage is synthesis speech module, which has functionality as follows:

- Analyze data input stream; sequence definite phonemes gets on the basis of text process,
- Supply instruments used to operate basis of diphones.
- Supply instruments to process digital signals.
- Execute concatenation of adequate signal samples.

Output of this module is acoustic speech signal. Synthesis speech module will be realized on the basis of concatenation. In this method performance based on already registered samples of natural speech signal.

3.2.3. Spoken language understanding system

An understanding and dialog component is required to manage interactions with the user and domain knowledge. It must be provided to guide the system's interpretation of speech and allow it to determine the appropriate action.

Spoken language understanding system for elementary speech input and output has a speech recognizer and a speech synthesizer, and a *sentence interpretation* component to parse the speech recognition results into semantic forms.

The central component that communicates with applications and the spoken language understanding modules such as discourse analysis, sentence interpretation, and response generation is the *Dialog Manager*. This component monitoring the flow of conversation tied to the action. It is also responsible for providing status needed for formulating responses, and maintaining the system's idea of the state of the discourse. The discourse information is crucial for sentence interpretation to interpret utterances in context.

4. Conclusions

There are a lot of advantages in applying language technology for preparing an interface between humans and computers. The main aim of integrating language technology with applications is to help disabled people and also to increase reliability, performance and user satisfaction while at the same time decreasing discomfort.

The example of such connecting is presented in this article voice controlled bank application. It is especially useful for disabled people. In the future this work could be expanded on connecting application with a biometric system using behavioral biometrics such as typing rhythm, and voice, which will be both identification and authentication.

References

- [1] Cernys P., Kubilius V., Macerauskas V., Ratkevicius K., *Intelligent Control of the Lift Model*, IEEE, 2003.
- [2] <http://www.zw.com.pl/arttykul/343320>.
- [3] Rabenja R., Amine S., Davrdson J., *Speech recognition for game control*, IEEE, 2004.
- [4] Balentine B., Morgan D.P., *How to Build a Speech Recognition Application: A Style Guide for Telephony Dialogues*, Enterprise Integration Group, April 1999.
- [5] Haleem M.S., *Voice Controlled Automation System*, IEEE, 2008.
- [6] Kuo J. *HMMS and Speech Recognition*, presentation.
- [7] Shirali-Shahrezal M., Shirali-Shahreza S., *CAPTCHA for Blind People*, IEEE, 2007.