Результати моделювання виявилися задовільними. Це підтверджує правильність побудованої моделі об'єкта прогнозування та вибору архітектури ШНМ для розв'язання задачі прогнозування технічного стану КС.

Висновок

В результаті досліджень відомих методів була запропонована нова модель ОП та відповідний метод прогнозування. Запропонований метод є універсальним, у тому розумінні, що уможливлює виконувати прогнозування стану як МПС в цілому, так і її окремих компонентів, причому в останньому випадку немає необхідності збирати всю інформація про систему, досить обмеженої кількості інформативних параметрів. Використання ШНМ як засобу прогнозування надає ряд переваг: врахування прихованих залежностей, можливість уточнення прогнозів за допомогою постійного донавчання мережі тощо.

1. Локазюк В.М., Медзатий Д.М. Збільшення тривалості функціонування обчислювальних систем за допомогою прогнозування: Матеріали Міжнародної науково-практичної конференції "Динаміка наукових досліджень". Том 1. "Сучасні комп'ютерні інформаційні технології". – Дніпропетровськ, 2002. – С. 25–27. 2. Касаев О.Б., Савченко В.И. Модели и методы прогнозирования технического состояния космических средств: Метод. пособие. – СПб., 1997. 3. Локазюк В.М., Медзатий Д.М. Концептуальна модель прогнозування технічного стану мікропроцесорних пристроїв на етапі експлуатації // Вісник технологічного університету Поділля. – 2003. – № 3. – С. 36–40. 4. Назаров А.В., Лоскутов А.И. Нейросетевые алгоритмы прогнозирования и оптимизации систем. – СПб., 2003. 5. Гаскаров Д.В., Голинкевич Т.А., Мозгалевский А.В. Прогнозирование технического состояния и надежности радиоэлектронной аппаратуры. – М., 1974. 6. Растригин Л.А. Пономарев Ю.П. Экстраполяционные методы проектирования и управления. – М., 1986. 7. Локазюк В.М., Поморова О.В., Медзатий Д.М. Метод прогнозування технічного стану комп'ютерних систем // Вісник Хмельницького національного університету. – 2005. – № 4. – Ч.1, Т.1 – С. 81–86.

J. Gadek
Department of Computer Science,
The College of Computer Science, Poland

THE DATABASE OF EMOTIONAL SPEECH

База даних емоційних розмов є частиною проекту "BaFra", метою якого є створення польсько мовного корпусу, що спеціалізується на тестуванні програм розпізнавання голосу. До складу бази входять фрази, продиктовані диктором яу у нормальній обстановці, так і модифіковані фрази. Подібна мовна база дозволяє тестувати стійкість алгоритмів розпізнавання мови до зміни диктора та його емоційного стану.

The database of emotional speech is the part of the „BaFra" project, witch has been designed in order to create the polish speech corpus specialized in testing voice recognition applications. The Corpus contains the phrases spoken by the speakers in a normal way as well as phrases spoken in a modified way. Such a language base permits to make algorithms resistance-tests which are responsible for speaker and emotion recognition

Introduction

Dealing with the speaker's emotion is one of the latest challenges in speech technologies. Three different aspects can be easily identified: speech recognition in the presence of emotional speech, synthesis of emotional speech, and emotion recognition. In this last case, the objective is to determine the emotional state of the speaker out of the speech samples. Possible applications include from help to psychiatric diagnosis to intelligent toys, and is a subject of recent but rapidly growing interest [1].

One of the major problems in the testing process of the speaker and speaker's voice recognition applications is lack of available, free of charge samples of voices recorded in different languages. The "BaFra" Project designed in The College of Computer Science in Lodz, Poland will cover this gap.

Under the supervision of the author of the article the structure of the database has been worked out containing speech samples spoken in polish language. It concerns mainly tests of speaker recognition applications although the structure had been prepared in such a way that it is possible to use it also to test applications detecting emotions in speech sample.

Corpus purpose

The main aim for developing this set was the preparation of speech database for speaker recognition applications tests, where the strong attention was focused on the resistance tests of such threats as: speaker's voice changes caused by emotions, passing time, illnesses or stupefactions; the speaker's conscious voice modification in order to mislead the testing application; the appearance of imitators that is persons pretended different people registered in the system, pretending somebody's else voice.

In this article, I am describing the first edition of set "BaFra" Corpus ver. 1.1. This version mainly contains a research material for testing the speaker recognition but its' underlying data construction and research material will permit to concerning emotions recognition in the speakers voice.

Corpus construction

Recording procedure

Sound recording was conducted by many recruited record-keepers chosen among the students of The College of Computer Science who were well trained for such a task. Each record-keeper had an identification number and was given a recording card where they could register the progress, parameters and the conditions of the recording process.

Record–keepers fond the speakers by themselves and recorded their voices. Then, they conducted at least five recording sessions with each speaker in certain intervals of time, where at least three sessions were conducted in the situation where the speaker had a changed voice. Unfortunately, in most cases those changes were conscious as the speakers tried to change their voices through the natural causes. After conducting the recording process, the record-keepers supplied the database operators with records and the recording cards so that the database operators could work out the records mainly by cutting up the sound files and inscribing adequate data into configuration files.

Equipment and file formats

All the records, according to the assumptions made during the creation of the project, were made on generally accessible equipment. In most cases it was a computer with a sound card and an average class microphone. The records were recorded in homesteads using different equipment with a different level of disturbances. All those parameters were registered in a recording card by the record-keepers, making a certain evaluation in certain cases following the instruction received during the course of training. For example, the conditions of recording (the noise factor) were evaluated at the scale from 1 to 10, where 10 means ideal conditions, unreachable in homesteads, 1 means major disturbances which make impossible to understand the text, whereas 7 – 8 stand for a quiet room, where the sound of a computer was the only noise. Such an approach permits to achieve such data, which can be obtained via Internet from an average user's disposal of a computer with a sound card and an average class microphone. All the records were implemented at the 44100 Hz sampling frequency, 16 bits, mono and registered in wav format files with PCM coding. Then, the whole research material was converted to 8000 Hz frequency (the telephonic lines quality). The whole research material is available in both versions.

Speakers

There are seventy speakers, whose voices are registered in the database, forty-eight women and twenty-two men. Four of them (two male and two female) was asked to simulate each of the six MPEG-4 defined emotional styles: anger, disgust, fear, joy, sadness and surprise plus a supplementary neutral one [2].

For each speaker in the configuration file there are registered only two information:

- Year of birth – on the basis of the date of birth and the date of a record it is possible to establish the age of a speaker at the moment of recording, there are voices of people at different ages in the database (at the time being, the youngest speaker is six years old and the oldest is seventy one),
- **Gender** – there is a field for marking gender in the database.

Subjective evaluation of emotional speech samples.

An informal subjective evaluation of emotional speech was carried out, with 2 non professional listeners according to [2]. All utterances were played to all the listeners, 50 for each emotion style, witch make 100 tests for one emotion style. Each listener had to choose between the seven emotional styles considered, valuing the intensity of the perception in an one to five scale. If the listener was not convinced of his choice, he could also mark a second one.

The results of the evaluation were quite satisfactory: more than an 80% of the first choices were correct, and this figure almost reaches 90% if second choices admitted. Besides, all errors was committed for short utterances. No notorious difference was observed in the results neither between one speaker and another nor between sessions.

Table 1

The confusion matrix of the emotion subjective evaluation experiment

	Surprise	Joy	Anger	Fear	Disgust	Sadness	Neutral
Surprise	86	6	8	0	0	0	0
Joy	0	90	8	0	0	0	0
Anger	6	4	82	0	0	0	0
Fear	7	0	1	94	5	7	0
Disgust	1	0	1	0	86	1	0
Sadness	0	0	0	6	6	91	0
Neutral	0	0	0	0	1	1	100
Total	86%	90%	82%	94%	86%	91%	100%

Table 1 shows the confusion matrix of the experiment. Columns represent the emotion elected in first choice for utterances belonging to the emotion of each row. The number of utterances per emotion is, in all cases, 100, with a grand total of 800. Errors mainly affect two different sets of emotions: first – fear, disgust and sadness; second – surprise, joy and anger. Most of the errors committed involved emotions of either of the two sets, which is usually confused with another emotion of the same set.

Phrases classification

The amount of phrases in the database depending on the age and the gender of speakers show the Table 2.

Table 2

Number of the registered phrases depending on age and gender of speakers

Age	Male	Female
less then 20	105	210
20-40	3675	1155
40-60	1050	735
more then 60	210	210
all	5040	2310

During each recording session the record-keepers recorded the following information in the configuration files:

- **The technical parameters of the record** – the sampling frequency (assuming 44100 Hz), the size of sampling (assuming 16 bites), the file format (assuming wav);
- **The equipment** – the description of equipment used for recording;
- **The evaluation of recording conditions** – the subjective evaluation of noise level in the scale from 1 to 10;
- **The speed of speech** – evaluation of speaking speed in a certain recording process in the scale from – 5 to +5, where 0 means the normal speed;
- **State of voice** – majority records represent the voices consciously modified by the speaker, the possible changes are: 1 - blocked nose, 2 - high tone of voice, 3 - low tone of voice, 4 - hoarseness, 5 - hard hoarseness, 6 - voice changed consciously by the speaker in order to make the recognition impossible, 7 - alien object in the mouth, 8 - difficulties in speaking caused by alcohol, 9 – whisper, 10 – hay fever, -1 – non-standard (according to the description placed in configuration files);
- **Emotional state of speaker** – which means: 0 – neutral, 1 – surprise, 2 – joy, 3 – anger, 4 – fear, 5 – disgust, 6 – sadness, unfortunately the classification based on emotions subjectively evaluated by record-keepers caused some difficulty, for example the state of disgust makes some problems for speakers.

For the convenience of searching for the adequate tasks sets of phrases the author of this research material implemented a division (classification) of the registered samples into classes and subclasses. Such a division has not been established forever. It is established that a future development is possible as well as the addition of new classes and subclasses, if needed.

There were four major classes distinguished of recorded phrases:

- **01 Class** – phrases common for all speakers of the database: the text repeated by each speaker during each recording session, phrases of this class permits to analyze changes of the same phrase depending on, for example, who speaks it, if it is a single speaker – they define his condition during the recording session;
- **02 Class** – phrases common only for a single speaker: the text repeated by each speaker during each recording session but not appearing for other speakers; phrases of this class permits to analyze how the same phrase changes depending on, for example, the condition of speaker during the recording session, how his voice is changing in the passing time taking under consideration the additional sound research material differing from the one of 01 Class; the content of the record in this class is established by the speaker itself or by the record-keeper while starting to record the first set;
- **03 Class** – optional phrases: the text changed during each recording session: it permits to test the recognition voice algorithms independent from the text: the content of the record in this class is established by the speaker himself or the record-keeper each time before the recording session;
- **04 Class** – imitating phrases: there are sampling of a certain speaker trying to pretend the phrase of a different subject, this class permits to test recognition algorithms resistance of voice for imitators pretending other people.

In each class we can distinguish the following subclasses:

- **001 Subclass** – sequences of figures spoken constantly: it permits to test the constant speech recognition for an average set of words or recognition of voice based on the sequence of figures,
- **002 Subclass** – sequences of figures spoken with intervals: it permits to test recognition of the isolated words for an average dictionary or recognition of voice based on very short phrases,
- **003 Subclass** – one-word phrases: it permits to test isolated words recognition from an average size dictionary or recognition of voice; the set of recommended words registered in 01 Class, what signifies each time and each speaker, was matched to contain all Polish phonemes (based on [1]);
- **004 Subclass** – a long text, it permits to test continuous speech recognition as well as phrases of freely matched length;

- **005 Subclass** – multi-words phrases, it permits to test continuous speech recognition as well as multi - words phrases treated as one unit, recognition of a subject based on established key-word; mainly in this example the phrases are the most valuable where the content had been established by the speaker himself, it imitates the situation where the subject himself invents the key-word for himself, securing the system against the unwanted access.

Particular classes are marked by class figures and subclasses figures, for example the class (03/005) determines 03 class and 005 subclass, that is optional multi-words phrases.

Described classes and subclasses can appear in each record or solely their subset. The minimal subset of the class is placed on the reverse side of the recording card and contains the following classes: (01/001), (01/002), (01/003), (01/004), (01/005), (02/005), and (03/005). Such a set, in the author's opinion, can be sufficient for characterizing a certain person's voice. In case where such a need appears the record-keeper was able to limit the amount of registered classes. For example, such a situation could take place when there was not appropriate amount of time to conduct a full record. Even a fragmentary record can enrich the researching material settled in the database. If needed, the record-keeper could also conduct a recording process containing all described classes, subclasses and what is more, a new yet non-existing class could be added. Such an activity has always been written off in the configuration database files.

Directory structure

After receiving the data from the record-keepers, the database operators, that are specialists responsible for introducing the data into the system, cut the records and placed them in adequate catalogues. For simplification of the above-mentioned activity I have invented (adopted) a simple inner database scheme, based on placing the registered records into a special catalogue structure. Descriptions are settled in the configuration files, written in the text format. Such an approach permits to introduce data into the database without the need of using any additional programming and without any additional specific skills, either. Any optional program with the option for cutting files in the "wav" format would be suitable, supplemented in a simple text editor. Such an approach permits to become independent from the used operating system. The database can be prepared on different systems and for the future purpose stored in a different operating system.

A basic structure of catalogues is described on the Figure 1. Inside each catalogue of the speaker there is a configuration file describing a concrete speaker. It consists of the gender, year of birth, identification number and any additional information. Each record-keeper has his four-digit identification number, given by the author of this research (for example: 0780). Each speaker has also his identification number, given by the record-keeper (for example: 0127). The record-keeper's identification number together with the speaker's identification number create eight-digit speaker's identification number, identifying a recorded speaker unequivocally (according to the previous examples it will be 0780-0127).

Each record has its' identification number given by the record-keeper, it consists of the date of the recording process showed in the two-numbers format – the year, the month, the day of week as well as the additional, two-digit number (for example 02112107 which stands for the record number 07 from the second of November, 2002).

The database structure catalogues are named in accordance with the adopted marks, Figure 2. The speaker's catalogue name consists of the recorder's identification and the speaker's identification. It consists of eight marks. The name of the catalogue record stands simply for the identification record number. Each recording catalogue consists of configuration file with the information about files recorded in an established format.

There are files containing registered phrases in the record catalogue. The filenames are always the same and reflect the phrase of the class according to my description from 3.4. section. For example, the file containing the phrase of the class (003/05) will receive the name "003 – 05.wav".

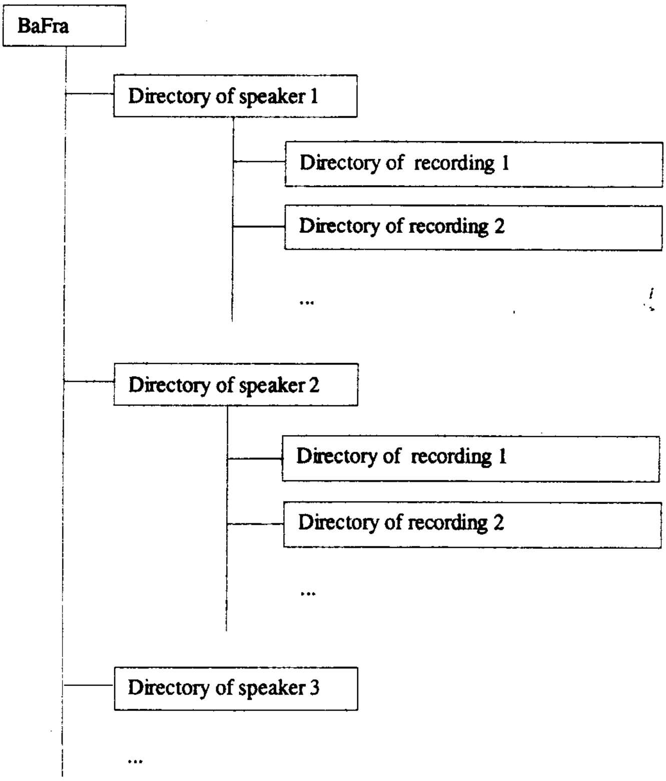
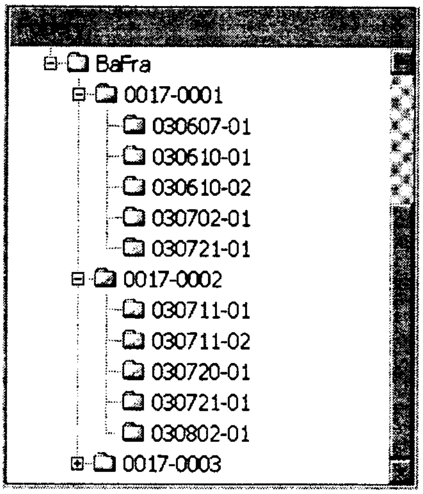


Figure 1: Directory structure



```
⊟ BaFra
  ⊟ 0017-0001
      030607-01
      030610-01
      030610-02
      030702-01
      030721-01
  ⊟ 0017-0002
      030711-01
      030711-02
      030720-01
      030721-01
      030802-01
  ⊞ 0017-0003
```

Figure 2: The fragment of real directory structure

Configuration files

The whole information concerning the files that contain the recorded phrases is placed in the configuration files. All the configuration text files remain in the same format. The files are divided into sections. The name of the section is signed in square brackets (e.g. [default] stands for default section) In each section there are datafields in a special format: field_name=value (see Figure 1, Figure 2. Everything that exists in a certain line after the semi-colon is treated as a comentary.

```
;user 0017-0003
reg_year=2003
born=1977
sex=m
note=
```

Figure 3: Example of "speaker.cfg" configuration file

In each speaker's catalogue there is a speaker's configuration file "speaker.cfg", Figure 3. In each catalogue of the record there is a configuration file of the record "rec.cfg", Figure 4. In the main database catalogue BaFra there is a configuration file "default.ctg" containing standard settings for all records. The structure of "default.cfg" file is the same as "rec.cfg" file. If there is not a description of a certain record in the "rec.cfg" file, it should remain in "default.cfg"

A good exapmle can be 01 class (according to the description in 3.4. section), which remains the same for each speaker and for that reason there is no need duplicate it in all configuration files.

The speaker's configuration file "speaker.cfg", Figure 3 contains only a few indispensable information for denoting the speaker, what signifies: the field "born" stands for the speaker's year of birth, the field "sex" stands for the speaker's gender, the field "reg_year" stands for the year of registration a certain speaker in the database and the field "note" stands for a commentary concerning the speaker. Additionally, there is a possibility of adding the speaker's identification number.

```
[default]
format=wav
rate=44100
bit=16
env=8
env_note="Silent room"
mic_note="Philips SBC MDI 10"
speed=0
speech=0
speech_note=Normal
emotion=0
emotion_note=Neutral
imit=0
[01-001.wav]
class=01/001
text="1234567890998448879632l2"
imit=1
imit01="0017-0001/030610-01/i.wav"
```

Figure 4: Part of the example "rec.cfg" configuration file

A configuration file of "rec.cfg" record, Figure 4, contains the information on the concrete record. [Default] section contains the information on the whole record, appearing in a certain catalogue – meaning – for all files appearing in this catalogue. If we wish placing the information concerning the concrete catalogue, we must create a section with the same name as the file. On the Figure 4, this is a section [01-001.wav] containing the information concerning the "01-001.wav" file.

In the file "rec.cfg" the fields represent the following meaning: the field "format" stands for the format of the file, the field "rate" stands for the sampling rate of the recording process; the field "bit" stands for the size of sampling; the field "env" stands for a record-keeper's subjective evaluation of the recording process conditions in the scale 1 – 10 (see section 3.5.); the field "env_note" consists of a written description of the recording conditions, the field "mic_note" consists of a written description of the used microphone; the field "speed" stands for the speech velocity, (see section 3.5.); the field "speech" stands for the speaker's state of voice, (see section 3.5.); the field "speech_note" contains a written description of the tone of the speaker's voice; the field "emotion" stands for the speaker's emotional state, (see section 3.5.); the field "emotion_note" contains a written description of the speaker's emotional state, the field "imit" possesses the value 0 or 1 and stands for the information whether there are imitation trials for certain files; the fields "imit01", "imit02",... accuar only when the "imit" field obtains the value 1 and contains the file names of the imitations; the field "text" contains a written description of the whole sound file

Discussion

A described database has been used for testing the voice recognition applications, witch has been made in the College of Computer Science. Those applications came into being using The Hidden Markov Model Toolkit (HTK) [4], pattern MASV – an experimental speaker verification system (published under the GNU GPL) [5]. In majority, the resistance of this application on threats described in section 2 was tested. There are plans to use it for emotion recognition. The results of the research will be presented in a separate paper.

Conclusions

In this paper I describe the first release of the "BaFra" corpus; "BaFra" version 1.1.The project has not been closed, yet; the corpus is still being developed, new phrases are still being added. The next release is being prepared at present as well as the user interface.

This corpus is freely available. The use and distribution are allowed under the GNU General Public License (an Open Source License, see://www.gnu.org).

Acknowledgements

Copyrights for this corpus belong to The College of Computer Science. The Department of Computer Science of The College of Computer Science financially supported this project. I thank students of The College of Computer Science for their help in preparing this corpus.

References

1. Cowie R, Douglas-Cowie E., Tsapatsoulis N., Votsis G., Kollias S., Fellenz W., Taylor J.G. Emotion recognition in human-computer interaction // IEEE Signal Processing magazine. – 2001. – vol. 18, no. 1. – PP. 32–80. 2. Albino Nogueiras, Asuncion Moreno, Antonio Bonafonte, and Jose B. Marino.Speech Emotion Recognition Using Hidden Markov Models // Proceedings of EUROSPEECH'01. – 2001. 3. Wisniewski M. Zarys fonetyki i fonologii wspolczesnego jezyka polskiego. – Nicolaus Copernicus University Press, Torun. – 2001. 4 Young S., Everman G., Kershaw D., Moore G., Odell J., Ollason D., Povey D., Valtchev V., Woodland P. The HTK Book.- Microsoft Corporation.- 1999. 5. Tuerk U., Schiel F. Speaker Verification Based on the German VeriDat Database // Proceedings of EUROSPEECH03. – 2003.