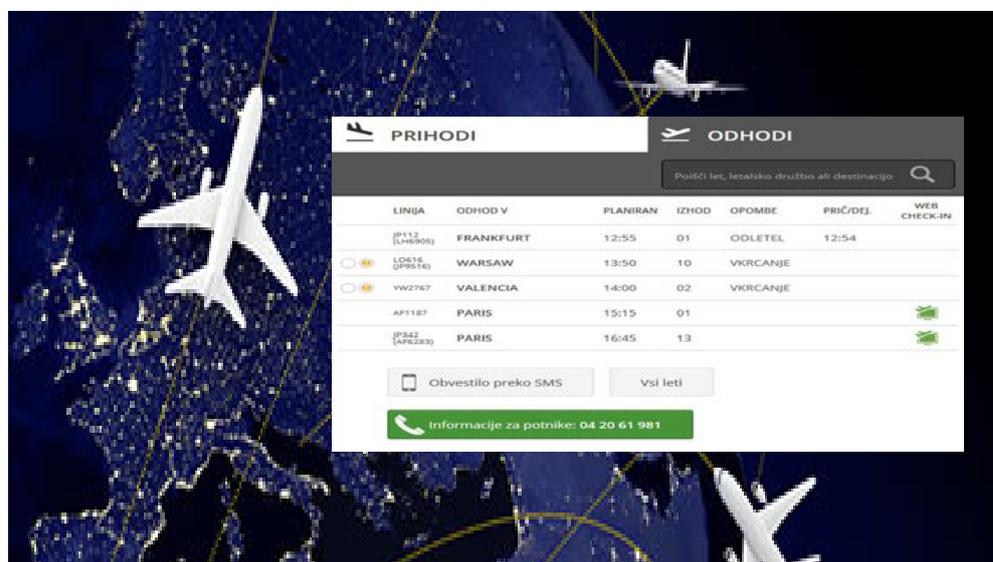


SLOVENIAN SPEECH DATABASE SOFES 1.0



The screenshot shows a flight information application interface. At the top, there are two tabs: 'PRIHODI' (Arrivals) and 'ODHODI' (Departures). Below the tabs is a search bar with the placeholder text 'Poišči let, letalsko družbo ali destinacijo'. The main content is a table of flight information. The table has columns for 'LINJA' (Line), 'ODHOD V' (Departure to), 'PLANIRAN' (Planned), 'IZHOD' (Departure), 'OPOMBE' (Remarks), 'PRIČ/DEJ' (Arrival/Departure), and 'WEB CHECK-IN'. The table lists several flights, including those to Frankfurt, Warsaw, Valencia, and Paris. Below the table, there are buttons for 'Obvestilo preko SMS' and 'Vsi leti'. At the bottom, there is a green button with a phone icon and the text 'Informacije za potnike: 04 20 61 981'.

LINJA	ODHOD V	PLANIRAN	IZHOD	OPOMBE	PRIČ/DEJ	WEB CHECK-IN
2112 (LH909)	FRANKFURT	12:55	01	ODLETTEL	12:54	
LD416 (UP916)	WARSAW	13:50	10	VKRCANJE		
YW2767	VALENCIA	14:00	02	VKRCANJE		
AF1187	PARIS	15:15	01			
JP342 (AF283)	PARIS	16:45	13			

June 2017

Speech Database of Spoken Flight Information Enquiries

This document provides a description of the database of spoken flight information enquiries in Slovene. The database is intended primarily for the development of acoustic speech models that are the key components of various speech technologies, such as speech recognizers and speech synthesizers, as well as biometric speaker-recognition systems, etc.

Simon Dobrišek, Jerneja Žganec Gros, Janez Žibert, France Mihelič, Nikola Pavešić et al.

University of Ljubljana, Faculty of Electrical Engineering
LUKS - Laboratory of Artificial Perception, Systems and Cybernetics
Tržaška 25, 1000 Ljubljana
simon.dobrisek@fe.uni-lj.si

Introduction

The SOFES speech database (Spoken Flight Enquiries in Slovene) is a collection of transcribed and segmented audio recordings of spoken flight-information enquiries in Slovene. SOFES is built on the basis of the GOPOLIS speech database [1, 2], which was acquired by the members of the Laboratory of Artificial Perception, Systems and Cybernetics at the Faculty of Electrical Engineering, University of Ljubljana in the period 1996–1998. The main purpose of the GOPOLIS speech database was the development of an automatic spoken-dialogue system for users who are enquiring about flight information over the telephone. The content of GOPOLIS is, however, sufficiently diverse to allow for the development of more generalized acoustic models of spoken Slovene, which are the key components of various speech technologies, such as speech recognizers and speech synthesizers, as well as biometric speaker-recognition systems, etc.

The audio recordings of the original GOPOLIS database are of two different qualities. Higher-quality recordings were obtained using a high-quality headset microphone and direct capturing of the voice signals through a computer's audio interface. Recordings of a lower quality were obtained using a traditional telephone handset and by capturing voice signals through a fixed telephone line. The recordings of two qualities make possible research on the impact of the quality of audio recordings on the performance of various speech technologies.

The original GOPOLIS database was extended in 2005–2008 with additional audio recordings of utterances that also relate to spoken flight-information enquiries. These additional recordings were acquired by students who attended the laboratory exercises of the course on Pattern Recognition, which was then given at the Faculty of Electrical Engineering, University of Ljubljana. The original database, which included voice recordings of 50 speakers (including 25 male speakers and 25 female speakers) who each uttered between 163 to 184 sentences, was extended with the additional voice recordings of 74 speakers (including 72 male speakers and 2 female speakers) who each uttered 20 additional longer sentences of spoken flight-information enquiries. These recordings were obtained under uncontrolled conditions and using different microphones and computer audio interfaces that were available to the students. The selection of the students was defined by their current enrolment in the given years of study. The sentences that were supposed to be uttered by the students were chosen at random from a large corpus of sentences that were automatically generated using a set of production rules, which were obtained on the basis of a detailed analysis of the actual live telephone dialogues between the users and the employees of the call centre of the Slovenian airline company, Adria Airways [3]

Besides the GOPOLIS database and the mentioned additional student voice recordings, the SOFES database also includes a less-extensive speech database K211D [4], which was acquired and edited in 1998 in order to carry out an analysis of certain phonetic units of spoken Slovene, where the emphasis was put on diphthongs. The K211D speech database includes the voice recordings of 10

speakers (including 5 male and 5 female speakers), who each uttered 251 carefully selected phonetically rich words.

The entire SOFES speech database contains the voice recordings of 134 speakers, of whom 102 are male and 32 are female speakers. The total number of all utterances is 12,536 and the total duration of all the microphone voice recordings is 9 hours and 52 minutes. The database also contains 8,546 telephone voice recordings of the same utterances that were simultaneously recorded via a microphone headset and are part the GOPOLIS database.

Additional information about the speakers who participated in obtaining the voice recordings, the time and place of the recording sessions as well as the equipment used can be obtained from the given references and literature as well as from specific files that are part of the SOFES database. In the rest of this document, some basic information about the structure and the content of the SOFES speech database is provided and a description of all the used file formats is given.

The structure of the SOFES database

The main content of the SOFES speech databases are, of course, audio recordings of speakers' utterances and their transcriptions. All the voice recordings are stored in the database in separate files that are located in the folder "utterances", which contains sub-directories for each of the speakers who are included in the database.

Speaker-identification codes

Each speaker is assigned a seven-character code that identifies one of the three original databases that are included in the SOFES database and from which the given speaker originates, as well as the speaker's serial number and gender. The seven-character speaker code is composed in accordance with the following template:

»dddd|ss|g«,

where the symbols denote the following:

»dddd«	Label of the original speech database, namely: »96sq«, »98kd«, »05rv«, »06rv«, »07rv« or »08rv«.
»ss«	Speaker's serial number, e.g.: »01«, »02«, ...
»g«	Speaker's gender, namely »f« (female) or »m« (male).

The database labels »05rv«, »06rv«, »07rv«, and »08rv«, respectively, denote the voice recordings that were obtained by the students who attended the laboratory exercises of the course on Pattern Recognition in 2005, 2006, 2007 and 2008. The same applies to the database labels »96sq" and "98kd", indicating 1996 and 1998 as the years in which the voice recordings of the GOPOLIS database ("96sq") and the KD211D database ("98kd") were obtained

If a certain file refers to any of the voice recordings that were obtained by the students who attended the laboratory exercises of the course on Pattern Recognition (no matter in which year) then the database label is given as »0xrv«.

In accordance with the above template, the speaker's code »98kd02m«, for example, indicates a male speaker with the serial number »02«, whose recordings were obtained in 1998 and originates from the K211D speech database. In the same manner, the speaker's code »08rv01f«, indicates a female speaker with serial number »01«, whose recordings were obtained in 2008 by a student who attended the laboratory exercises of the course on Pattern Recognition.

Voice-recording files

Voice recordings of all the utterances are stored in the usual file format WAVE/WAV (Waveform Audio File Format), which is based on the RIFF file format (Resource Interchange File Format).

They were captured in mono, with a sampling frequency of 16 kHz, and using a 16-bit linear quantization of the audio signal amplitudes.

The name of each voice-recording file with the extension ".wav" is designed as a 12-character code that is composed in accordance with the following template:

»ssssss|nnnn|c«.wav,

where the symbols denote the following:

»ssssss«	Speaker code, as described in the previous sub-section, e.g.: »96sq01m«, »98kd02f«, »05rv03m«, ...
»nnnn«	Four-digit utterance identification number, e.g.: »0001«, »0002«, »0003«, ...
»c«	Recording channel, namely »m« (microphone) or »t« (telephone).

In accordance with the above template, the file name »98kd02m0001m.wav«, for example, indicates a voice recording of a male speaker with the speaker identification code »98kd02m«, who uttered a sentence with the identification number »0001« that originates from the K211D speech database, where the recording was captured using a headset microphone. In the same manner, the file name »96sq11m0001t.wav« indicates a voice recording of a male speaker with the speaker identification code »96sq11m«, who uttered a sentence with the identification number »0001« that originates from the GOPOLIS speech database, where the recording was captured using a telephone handset.

Transcriptions of the voice recordings

All the voice recordings of the SOFES database are accompanied by orthographic and phonemic transcriptions with time boundaries. Different types of transcriptions and segmentations are provided in separate files with the same name as the voice-recording files but with different extensions. The contents of the folder »utterances« as illustrated by the following structure:

```

└─ utterances
   └─ 05rv01f
      └─ 05rv01f0001m.pho
      └─ 05rv01f0001m.trs
      └─ 05rv01f0001m.txt
      └─ 05rv01f0001m.wav
   ...
   └─ 96sq01f
      └─ 96sq01f0001m.pho
      └─ 96sq01f0001m.trs
      └─ 96sq01f0001m.txt
      └─ 96sq01f0001m.wav
   ...
   └─ 98kd01f
      └─ 98kd01f0001m.pha
      └─ 98kd01f0001m.phe
      └─ 98kd01f0001m.pho
      └─ 98kd01f0001m.trs
      └─ 98kd01f0001m.txt
      └─ 98kd01f0001m.wav
   ...

```

The filename extensions determine the format and the content of the transcription files that are associated with the audio files with the extension ".wav". All the orthographic transcriptions are stored in the UTF-8 character encoding (<http://www.utf-8.com>), and all the phonemic transcriptions are encoded using the SAMPA phonetic alphabet (<http://www.phon.ucl.ac.uk/home/sampa>) that consists of a mapping of symbols of the International Phonetic Alphabet onto ASCII codes.

The filename extensions define the following format and content of the transcriptions.

».txt«	<p>A plain-text transcription of the corresponding voice recording in a single line with the tag »<s>« at the beginning and the tag »</s>« at the end of the transcription. The transcription is provided as a plain sequence of words, separated by space characters, where no punctuation or capitalization is used.</p> <p>For example, the text transcription of the voice recording »05rv23m0001m.wav« is provided in the plain text file »05rv23m0001m.txt«, which has the following content:</p> <pre><s> ali je res objavljen odhod airfranceovega letala v genovo v torek tega meseca pozno zvečer ob štiriindvajset minut čez osem </s></pre>
».trs«	<p>A text transcription and segmentation of the corresponding voice recording in an XML file format that is supported by a computer program for annotating, transcribing and segmenting speech signals - »Transciber« (http://trans.sourceforge.net).</p>
».pho«	<p>Broad-phonetic transcription and segmentation of the corresponding voice recording. Broad-phonetic transcriptions are based on the use of a set of 32 phonetic units, representing the basic phonetic variants (allophones) of the phonemes of spoken Slovene, as well as the special unit »sil« that represents pauses in the speech [4].</p> <p>The list of phonetic units used in the broad-phonetic transcriptions is provided in the folder »annotations« in the file »annotations/sofes-pho.lst«.</p> <p>For example, the broad-phonetic transcription of the voice recording »98kd01m0029m.wav« is provided in the plain text file »98kd01m0029m.pho«, which has the following content:</p> <pre>0 3056 sil 3056 3824 d 3824 4496 Z 4496 5872 u 5872 7568 n 7568 8016 g 8016 9008 l 9008 11536 a 11536 14816 sil</pre> <p>The above numbers are the beginning and the ending sample numbers of the intervals within which the phonetic units that are given in the same rows are articulated in the corresponding voice recording. Due to the fact that the voice recordings were captured using a sampling frequency of 16 kHz, for example, the value of 3056 indicates a point in time 0.191 s (191 ms) from the beginning of the voice recording.</p>
».phe«	<p>Narrow-phonetic transcription and segmentation of the corresponding voice recording. Narrow-phonetic transcriptions are only provided for the voice recordings that originate from the K211D speech database and are intended for more specific and detailed phonetic studies of spoken Slovene.</p> <p>Narrow-phonetic transcriptions are based on the use of a set of 46 phonetic units, representing the basic phonetic variants of the phonemes and the long and short variants of the Slovene vowels, as well as the special units »sis« and »sie« that represent initial and final pauses in speech [4].</p>

	<p>The list of phonetic units used in the narrow-phonetic transcriptions is provided in the folder »annotations« in the file »annotations/sofes-pho-98kd-phe.lst«.</p> <p>For example, the narrow-phonetic transcription of the voice recording »98kd01m0029m.wav« is provided in the plain text file »98kd01m0029m.phe«, which has the following content:</p> <pre>0 3056 sis 3056 4496 dZ 4496 5872 u: 5872 7568 N 7568 8016 g 8016 9008 l 9008 11536 a 11536 14816 sie</pre> <p>The above numbers indicate the beginning and the ending sample numbers of the interval within which the given phonetic units are articulated in the corresponding voice recording, as already described for the broad phonetic transcriptions.</p>
».pha«	<p>Acoustic-phonetic transcription and segmentation of the corresponding voice recording. Acoustic-phonetic transcriptions are only provided for the voice recordings that originate from the K211D speech database and are intended for more specific and detailed acoustic-phonetic studies of spoken Slovene.</p> <p>Acoustic-phonetic transcriptions are based on the use of a set of 58 phonetic units, including all the phonetic units that are used for narrow phonetic transcriptions as well as the additional phonetic units representing the closure, obstruction and aspiration phases of plosives and affricates, and the special units »sis« and »sie« that represent the initial and final pauses in the speech [4].</p> <p>The list of phonetic units used in the acoustic-phonetic transcriptions is provided in the folder »annotations« in the file »annotations/sofes-pho-98kd-pha.lst«.</p> <p>For example, the acoustic-phonetic transcription of the voice recording »98kd01m0029m.wav« is provided in the plain text file »98kd01m0029m.pha«, which has the following content:</p> <pre>0 3056 sis 3056 3600 dZc1 3600 3824 dZp1 3824 4496 dZ 4496 5872 u: 5872 7568 N 7568 7856 gc1 7856 8016 g 8016 9008 l 9008 11536 a 11536 14816 sie</pre> <p>The above numbers indicate the beginning and the ending sample numbers of the interval within which the given phonetic units are articulated in the corresponding voice recording, as already described for the broad phonetic transcriptions.</p>

Assembled transcriptions of all the voice recordings

In addition to the above-mentioned lists of phonetic units, the subdirectory »annotations« contains several variants of a pronunciation dictionary for all the words that appear in the speech transcriptions, an extensive corpus of automatically generated sentences that relate to the spoken flight-information enquiries in Slovene, as well as several assembled files in XML format that contain descriptions, transcriptions and segmentations of all the voice recordings that are included in the SOFES speech database.

The sub-directory »annotations« contains the following files:

```
— annotations
  — sofes-lex-sampa.txt
  — sofes-lex-w3c-sampa.xml
  — sofes-lex-w3c.xml
  — sofes-pho-98kd-pha.lst
  — sofes-pho-98kd-phe.lst
  — sofes-pho.lst
  — sofes-utt-0xrv-tei.xml
  — sofes-utt-96sq-tei.xml
  — sofes-utt-98kd-tei.xml
  — sofes-utt-gen.txt
  — sofes-utt.txt
```

The following is a brief description of the content and format of the above files.

»sofes-lex-sampa.txt« »sofes-lex-w3c-sampa.xml« »sofes-lex-w3c.xml«	<p>Three variants of a pronunciation dictionary for all the words that appear in the transcriptions of the voice recordings of the SOFES speech database. Word pronunciations are given using the phonetic units that are used for the broad-phonetic transcriptions of voice recordings. The same pronunciation dictionary is given in three different file formats.</p> <p>The file »sofes-lex-sampa.txt« contains the pronunciation dictionary in a plain text format, where each text line contains an orthographic representation of a given word, followed by its broad-phonetic transcription separated by one or more space characters. The X-SAMPA phonetic alphabet (http://www.phon.ucl.ac.uk/home/sampa) is used for encoding the phonetic units that appear in the pronunciations.</p> <p>For example, a part of the contents of the file "sofes-lex-sampa.txt" looks as follows:</p> <pre>... airwaysa E r w E I s a airways E r w E I s airwaysom E r w E I s O m airwaysovega E r w E I s O v E g a airwaysov E r w E I s O U airwaysovih E r w E I s O v i x airwaysovo E r w E I s O v O čakaj tS a k a I alenska a l e n k a ali a l i alojz a l O I s ...</pre> <p>The file »sofes-lex-w3c-sampa.xml« contains the pronunciation dictionary in a W3C XML file format (http://www.w3.org/TR/pronunciation-lexicon), where the mentioned X-SAMPA phonetic alphabet is used for encoding word pronunciations.</p>
---	---

	<p>Parts of the contents of the file »sofes-lex-w3c-sampa.xml« look as follows:</p> <pre><?xml version="1.0" encoding="UTF-8"?> <lexicon version="1.0" ... alphabet="x-sampa" xml:lang="sl-SI"> ... <lexeme> <grapheme>airwaysovo</grapheme> <phoneme>E r w E I s O v O</phoneme> </lexeme> <lexeme> <grapheme>čakaj</grapheme> <phoneme>tS a k a I</phoneme> </lexeme> <lexeme> <grapheme>alenska</grapheme> <phoneme>a l e n k a</phoneme> </lexeme> ...</pre> <p>The file »sofes-lex-w3c.xml« contains the pronunciation dictionary in the above-mentioned W3C XML file format, where the IPA phonetic alphabet is used for encoding the word pronunciations (http://www.internationalphoneticalphabet.org).</p> <p>Parts of the contents of the file »sofes-lex-w3c.xml« look as follows:</p> <pre><?xml version="1.0" encoding="UTF-8"?> <lexicon version="1.0" ... alphabet="ipa" xml:lang="sl-SI"> ... <lexeme> <grapheme>airwaysovo</grapheme> <phoneme>ɛ r w ɛ i s ɔ v ɔ</phoneme> </lexeme> <lexeme> <grapheme>čakaj</grapheme> <phoneme>tʃ a k a ɪ</phoneme> </lexeme> <lexeme> <grapheme>alenska</grapheme> <phoneme>a l e n k a</phoneme> ...</pre>
<p>»sofes-pho.lst« »sofes-pho-98kd-phe.lst« »sofes-pho-98kd-pha.lst«</p>	<p>Plain lists of phonetic units that are used for broad-phonetic (»sofes-pho.lst«), arrow-phonetic (»sofes-pho-98kd-phe.lst«) as well as acoustic-phonetic (»sofes-pho-98kd-pha.lst«) transcriptions of voice recordings, as described in the previous section. All the phonetic units in these lists are encoded using the SAMPA phonetic alphabet (http://www.phon.ucl.ac.uk/home/sampa).</p>
<p>»sofes-utt.txt«</p>	<p>Plain text transcriptions of the voice recordings, where each text line contains one such transcription with the tag »<s>« at the beginning and the tag »</s>« at the end of the transcription. The transcriptions are provided as plain sequences of words separated by space characters, where no punctuation or capitalization is used. At the end of each text line, an identification code of the voice recording that refers to the given transcription is given in brackets. The design of the voice-recording identification codes is described in the previous subsections. This file contains a total of 21,082 transcriptions of voice recordings in as many text lines.</p>

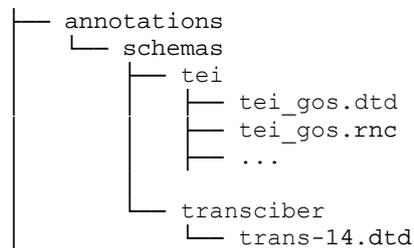
	<p>A part of the content of the file »sofes-utt.txt« looks as follows:</p> <pre> ... <s> halo </s> (96sq20m0001m) <s> ja </s> (96sq20m0002m) <s> je kdo tam </s> (96sq20m0003m) <s> se slišimo </s> (96sq20m0004m) <s> nič ne slišim </s> (96sq20m0005m) <s> kaj je zdaj to </s> (96sq20m0006m) ... </pre>
»sofes-utt-gen.txt«	<p>Plain text transcriptions of a large number of automatically generated sentences [3]. These sentences had not been actually uttered by the speakers. They are meant to be used for the statistical (language) modelling of the spoken language that is used in spoken flight-information enquiries in Slovene.</p> <p>Automatically generated sentences are given in single plain text lines with the tag »<s>« at the beginning and the tag »</s>« at the end of each sentence. The sentences are provided as plain sequences of words separated by space characters, where no punctuation or capitalization is used. This file contains a total of 1,025,164 sentences in as many text lines.</p> <p>A part of the contents of the file »sofes-utt-gen.txt« looks as follows:</p> <pre> ... <s> ali prileti letalo iz gran canarie </s> <s> obstaja let s prestopom v telaviv </s> <s> morda leti kakšno letalo iz frankfurta </s> <s> koliko stane karta leipzig larnaca </s> <s> katere dni v tednu letite iz helsinki </s> ... </pre>
<p>»sofes-utt-96sq-tei.xml« »sofes-utt-0xrv-tei.xml« »sofes-utt-98kd-tei.xml«</p>	<p>Descriptions, transcriptions and segmentations of voice recordings in a unified XML format that is meant to be used for building spoken-language databases, and was proposed by the TEI initiative (http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TS.html).</p> <p>Each of the three files refers to one of the three parts of the SOFES database that originate from one of the three original speech databases, where the label "96sq" refers to the original GOPOLIS database, the label "98kd" refers to the original K211D database, and the label "0xrv" refers to the original database of voice recordings that were obtained by the students in the period 2005 to 2008, as explained in the introductory section.</p> <p>In addition to the text transcriptions of all the voice recordings, these three files also contain some basic information on the speakers, such as their age and education, as well as some basic information on the time and place of the recording sessions and the equipment used.</p>

The main components of the SOFES speech database are therefore located in the file folders »utterances« and »annotations«. The remaining contents are primarily some basic examples about how to use the presented speech database. The most important such example is the proposed division of the database into training, development and test datasets. Compliance with the

proposed protocol makes possible reliable comparisons of different approaches to acoustical and language modelling of speech that is based on the use of the SOFES database.

XML document type definitions

The sub-directory »annotations/schemas« contains the following files



The files with the extension ».dtd« (Document Type Definition) define the structure, the elements and the attributes of the xml files that contain the descriptions, transcriptions and segmentations of the SOFES voice recordings and are described in the previous sub-sections.

The sub-directory »annotations/schemas/tei« contains the file »tei_gos.dtd« and the other files that define the structure, the elements and the attributes of the XML files »sofes-utt-96sq-tei.xml«, »sofes-utt-0xrv-tei.xml« and »sofes-utt-98kd-tei.xml« that are described in above. The three files contain the descriptions, transcriptions and segmentations of the SOFES voice recordings in the XML format that is meant to be used for building spoken-language databases, and was proposed by the TEI initiative (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TS.html>) with the aim of supporting better interchangeability of different databases between the research groups working in this field of research.

The sub-directory »annotations/schemas/transcriber« contains the file »trans-14.dtd« that defines the structure, the elements and the attributes of the XML files with the extension ».trs« that are created and used by the »Transcriber« toolkit (<http://trans.sourceforge.net>), a computer programme for editing, annotating, transcribing, and segmenting voice recordings.

Database usage examples

The SOFES speech database is primarily intended for different phonological and linguistic analyses of spoken Slovene, as well as for the development of computerized systems for automatic speech recognition, speaker recognition, and speech synthesis. These systems are mainly based on the use of various methods of automatic pattern recognition and machine learning, and an evaluation of these methods requires the availability of training, development and test datasets.

In the file folder »examples«, several example files are available that can be used for such research and development activities.

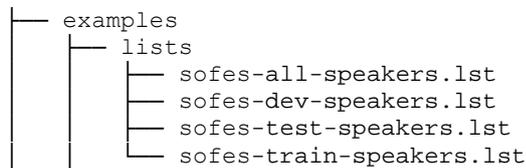
The lists of training, development and test speakers

As explained in the introduction section, the voice recordings of the SOFES database are divided into three parts that relate to the three original speech databases that are included in the database. This division is, however, not suitable to be used for creating the training, development and test datasets of voice recordings for machine-learning and pattern-recognition methods, as they are not appropriately balanced. For instance, the K211D speech database contains the utterances of the phonetically balanced isolated words that do not necessarily relate to the spoken flight-information enquiries. On the other hand, the voice recordings that were obtained by the students contain only longer utterances, and in addition to that, the selection of speakers was not gender balanced (male speakers dominate over female speakers).

For these reasons, an additional, completely random, division of the SOFES database into three parts was carried out. The three parts are balanced in terms of speaker gender as well as in terms of the content of the selected utterances. The division of the voice recordings is actually based on

the division of the speakers. This allows for the creation and evaluation of speech systems that are speaker-independent.

The division of speakers into the training, development and test datasets is provided in the following files with the extension ».lst«



These files contain plain lists of the identification codes of the speakers whose voice recordings are included in the training (»-train-«), the development (»-dev-«) or the test (»-test-«) dataset. This division allows for the systematic development, evaluation and comparison of automatic pattern-recognition and machine-learning methods that are being developed and used in the field of speech technologies (e.g., speech recognizers, speaker recognizers, speaker gender or age recognizers, etc.).

The three sets of speakers include voice recordings that, in total, comprise:

»sofes-train-speakers.lst«	Voice recordings of 75 speakers of a total duration of 7 hours 2 minutes in 41 seconds.
»sofes-dev-speakers.lst«	Voice recordings of 27 speakers of a total duration of 1 hour 19 minutes in 34 seconds.
»sofes-test-speakers.lst«	Voice recordings of 32 speakers of a total duration of 1 hour 29 minutes in 44 seconds.

The file »sofes-all-speakers.lst« comprises the identification codes of all the speakers who are included in the SOFES speech database.

Conclusions

According to the present criteria, the SOFES speech database is not very extensive. However, it contains enough data to support additional basic and educational research and development of speech technologies for spoken Slovene. Additional information about the original speech databases, which are included in the SOFES database, can be obtained from the references below or directly from the authors.

References and literature

- [1] S. Dobrišek, J. Žganec Gros, I. Ipšič, K. Pepelnjak, F. Mihelič and N. Pavešič, "GOPOLIS: Slovenska podatkovna zbirka govornih poizvedovanj," in *Zbornik konference: Jezikovne tehnologije za slovenski jezik*, Ljubljana, 1998.
- [2] F. Mihelič, J. Gros, S. Dobrišek, J. Žibert and N. Pavešič, "Spoken language resources at LUKS of the University of Ljubljana," *International Journal of Speech Technology*, vol. 6, no. 3, pp. 221-232, 2003.
- [3] S. Dobrišek, J. Žganec Gros, F. Mihelič and N. Pavešič, "Recording and labelling of the GOPOLIS Slovenian speech database.," in *Proceedings, First International Conference on Language Resources & Evaluation*, Granada, Spain, 1998.
- [4] S. Dobrišek, *Analysis and Recognition of Phones in Speech Signal* : PhD Thesis (In Slovenian), Ljubljana: University of Ljubljana, Faculty of Electrical Engineering, 2001.