JURISDIC – Polish Speech Database for taking dictation of legal texts

Grażyna Demenko[1], Stefan Grocholewski[2], Katarzyna Klessa[1], Jerzy Ogórkiewicz[3], Agnieszka Wagner[1], Marek Lange[3], Daniel Śledziński[1], Natalia Cylwik[1]

[1]Institute of Linguistics, Adam Mickiewicz University, Poznań

[2]Institute of Computing Science, Poznań University of Technology

[3]Laboratory of Speech and Language Technology , Adam Mickiewicz University Foundation, Poznań

E-mail: lin@amu.edu.pl, stefan.grocholewski@cs.put.poznan.pl, klessa@amu.edu.pl, sova_jo@sylaba.poznan.pl, wagner@amu.edu.pl, marek.lange@gmail.com, danielsl@poczta.onet.pl, nataliac@amu.edu.pl

Abstract

The paper provides an overview of the Polish Speech Database for taking dictation of legal texts, created for the purpose of LVCSR system for Polish. It presents background information about the design of the database and the requirements coming from its future uses. The applied method of the text corpora construction is presented as well as the database structure and recording scenarios. The most important details on the recording conditions and equipment are specified, followed by the description of the assessment methodology of recording quality, and the annotation specification and evaluation. Additionally, the paper contains current statistics from the database and the information about both the ongoing and planned stages of the database development process.

1. Introduction

Current speech recognition systems rely heavily on databases whose size and structure depend more or less on their particular application. As evaluation of the current ASR systems shows (J. Loof et. al., 2007, Docio-Fernandez, 2006) it is necessary to create appropriate speech databases which would take into account as many sources of speech variability as possible (Gibbon et. al., 1997). Database specification and validation of ASR systems for 20 European languages have been lately carefully verified within the SPEECON project. Also, a great effort has been made to evaluate various speech databases for SLT systems within the TC-STAR project. The inspection of the collection of ELRA Language Resources enables the assessment of existing European databases for different applications and languages.

The aim of the JURISDIC project is to create a database for the needs of taking dictation of legal texts. A review of the results of ASR systems developed for other languages shows that while creating such a system for Polish there is a need to modify some assumptions concerning acoustic-phonetic database structure. Some problems are universal like adequate coverage of segmental and suprasegmental structure, others however are connected with language-specific features (e.g. ensuring a full coverage of Polish consonant clusters in the speech database).

The general assumptions for the Polish JURISDIC database take into account the acoustic, phonetic and grammatical factors, some of which can be controlled, at least to some extent, in a prepared, fixed part of the database. As regards semantic structure, it depends strongly on the situational context and thus in case of JURISDIC database only (semi)spontaneous using of authentic legal texts and police reports dictation can guarantee appropriate semantic coverage.

2. The Structure of JURISDIC Database

The variable part of the database will include speech delivered by 1000 speakers. The recordings included in the corpora come from: a) the court (speech by a judge), b) the legal/notary's/prosecutor's office (speech by a lawyer), c) the police station (speech by a police officer), approx. 500 voices, d) office/university: approx. 300 voices. The distribution of sex and age is approximately 50:50. Although Polish is not very diverse as far as dialects are concerned, the recordings have been done in 16 main districts of Poland. The session recorded for each speaker consists of approximately 20-40 min of semi-spontaneous speech and, depending on the speech tempo, approximately 30 min of read speech (about 170 shorter and longer sentences). The speakers are asked to read a text as in a dictation task. Table 1 below shows JURISDIC speech corpus contents.

A. Semi Spontaneous Speech

Sub-corpus 1A. Spontaneous Dictation (legal, police, court vocabulary)

This sub-corpus contains formal speech (dictation on various application topics). Typical tasks are: dictation of any kind of legal texts (areas: judicial, disciplinary, criminal, divorce) in court, police reports (different topics, e.g. a description of a theft, burglary using common vocabulary, etc.). The number of the recorded topics varies between speakers.

Sub-corpus 2A. Spontaneous Dictation (common topics)

This sub-corpus contains informal speech (dictation on various common topics). Typical tasks are: a description of a birthday, giving directions, giving an excuse, a description of holidays, etc. The speaker is requested to be speak in a neutral style following instructions such as: *Imagine that you are calling your friend/father/boss and telling them something/excusing yourself/deciding on something*, etc. The number of the recorded topics varies between speakers.

Sub-corpus 3A. Elicited Dictation (Answering questions)

The aim of sub-corpus 3A is to obtain some semantically important, frequent items such as birth dates, relative dates, times of day, city names, proper names, age, money amounts, currencies, sequences of digits and numbers, telephone numbers, mathematical operations as well as answers like yes/no/maybe, etc. and education, profession, etc. (27 categories).

B. Read Speech. Grammatically and Phonetically Controlled Structure

Sub-corpus 1B. Phonetically controlled structure. Syntactically complex sentences.

By ‘syntactically complex’ we mean: a) variable concatenation of phrases, b) variable phrase length. By ‘phonetically controlled’ we mean: adequate coverage of triphones, triphones in the final position of a word/phrase. For selection of the phonetically rich sentences (from 3000 sentences) the following constraints are set: each speaker produces 60 complex sentences, each sentence is read by 15-20 speakers.

Sub-corpus 2B. Phonetically controlled structure. Syntactically simple sentences

We expect that 90 short sentences will be provided by each speaker with the explicit intention of obtaining an adequate coverage for the chosen consonant clusters, short bigrams and triphones both in the accented and unaccented position. The whole 2B Corpus should contain approx. 4000 sentences.

Each sentence should be read by 20 speakers. The main aim of the Corpus B was to obtain:

a) CVC triphones in context of sonorants in a chosen accented/unaccented position. The number of accented positions depends on a particular word’s frequency, e.g. for triphone: jem (I eat/I am eating) we have 4 prosodic positions e.g. *Łososia dziś jemy*? (Eng. *Are we eating salmon today*?).

b) CVC triphones in context of voiced consonants in a chosen accented/unaccented position. The number of accented positions depends on a particular word’s frequency. The whole subdatabase has approx. 800 sentences with controlled consonant clusters. The voiced context for the accented triphones was chosen because of a strong influence of accent on acoustic features of the triphone (especially the sonorant-vowel connection is extremely context dependent).

c) Examples of short bigrams in utterance initial position. The whole sub-database consists of approx. 2000 sentences with the controlled bigrams (e.g. two conjunctions, conjunction and preposition, etc.) in initial position and in the middle of a phrase for the most frequent bigrams. The short (one- or two-syllable) words are most difficult to recognize for ASR systems. Table 2 shows some examples of bigrams. The absolute frequency of different bigrams in Polish is given in brackets (based on the analysis of twenty million words taken from newspaper texts). frequency of different bigrams in Polish is given in brackets (based on the analysis of twenty million words taken from newspaper texts).

Corpus	Sub-corpus	Duration	Description (number of items per speaker)
A. Semi Spontaneous, Elicited, Descriptive, Controlled Dictation	1A 2A	20-40min	Free semi-spontaneous speech (dictation on various application topics). Free semi spontaneous speech (dictation on common topics).
	3A	3 min	Elicited spontaneous speech (answering questions, etc.). 27 questions.
B.Read speech. Grammatically and phonetically controlled structure	1B 2B 3B	20 min	Grammatically and phonetically controlled structure 1.Syntactically complex sentences – 60. 2.Syntactically simple sentences – 90. 3.Special lexical phrases (words) – 7.
C. Read speech. Core words and application phrases, texts	1C 2C	10 min 10 min	Semantically controlled structure 1.General purpose words and phrases 2.Application-specific short texts for users' needs

Table 1. Corpus content definitions

Bigram	Frequency	Phrase
i w	(7127)	***I w** piątek też się widzieliśmy.* (lit. ***And on** Friday we saw each other as well*)
a w	(5012)	***A w** sobotę idziemy do kina.* (lit. ***And on Saturday** we are going to the movies*)
i z	(2422)	***I z** tobą też muszę porozmawiać* (lit. ***And with** you I also need to speak*)

Table 2: Examples of Polish bigrams based on a statistical analysis.

d) Examples of consonant clusters: the whole sub-database consists of approx. 800 sentences with controlled consonant clusters. Special attention was given to CCCC and CCCCC clusters like: pstf, mpstf: głupstwo, skąpstwo (Eng. nonsense (or trifle), avarice).

Sub-corpus 3B. Special lexical phrases (words) The sub-corpus with more than 400 short one- or two-word includes special words like modulants, greetings, jargon/vulgar expressions. It was constructed manually based on dictionaries and other resources for Polish. At least 7 items are provided by one speaker.

Triphone statistics

The overall statistics of triphone coverage within the whole B corpus is as follows: triphones within word: 10593, triphones containing an accented vowel: 8492, unaccented triphones 10650, triphones in phrase final position: 4495.

Triphone lists serving as reference for the purpose of manual preparation of the B text corpus were created as follows: 2 million words were randomly selected from a corpus of texts including about 10 million words. This selection was automatically transcribed using modified SAMPA notation. An inventory of 39 phonemes was assumed. Syllable boundaries and accent annotation was based on rules proposed by Demenko et al., 2003. On the basis of the two-million-word set the list of all triphones found in this set was produced. Besides, the list included the information of the number of occurrences within the two million set and the list of words containing the respective triphone, only the triphones occurring within words (and not across word boundaries) were taken into account. The list do not deliver all possible Polish triphones, however it was assumed that if a triphone was not found in a randomly selected two-million-word set, it may be regarded as a very rare triphone and thus omitted.

C. Read Speech. Semantically Controlled Structure

Sub-corpus 1C. General purpose words and phrases Within this group utterances are divided into: general words/phrases and general-purpose commands. The general-purpose words/phrases include 33 categories, among them: isolated digits, numerals, measures, letters, special keyboard characters, special legal acronyms, emails, web addresses. No instructions are given to speakers as to how to spell these items.

Sub-corpus 2C. Application-specific short texts for users' needs

Texts extracted from original police reports and professional legal documents (up to 100 sentences).

3. Recording Conditions and Equipment

3.1 Recording Environment

Creating a large voice database is a great logistic task and requires specific recording equipment (both hardware and software). For the purpose of the present project office environment was assumed to be the target environment. A standard office is a relatively quiet area where the stationary background noise characteristics is close to white noise. Reverberation is on low or medium level. It was decided to obtain stereo recordings from two microphone positions: a 'close distance' and 'medium distance' position using a headset microphone and a 'table' microphone. Both of them are electret microphones with cardioid characteristics (typical low-budget devices). Headset microphone is mounted close to the speaker's mouth and the acquired recordings are expected to be clean, i.e. with good signal-to-noise ratio and very low reverberation. But the level of 'pops' and 'breathing' noises can be relatively high depending on microphone position. The table microphone is regarded as 'speaker-independent' (distance from the speaker's mouth to table microphone is approx. 0.5m) but signal-to-noise ratio is lower and reverberation level is higher. Due to the emphasis of low frequencies of the directional types in a near field, the frequency characteristics of 'close distance' recording channel might be compensated by using a specialized microphone (e.g. Sennheiser ME104) or by high pass filtering. But because of commonness of this phenomenon in almost all available microphones, the compensation can be abandoned.

3.2 Hardware

Two types of microphones are used: Sennheiser ME-3 for 'close distance' position (delivered as part of a wireless system used at the beginning of the project, e.g. for one-channel recordings of judges in courtrooms), and AKG C-1000S – for the 'middle distance'. Finding the proper analog to digital converter appeared to be a problem to a certain extent. Most of them are simple, mono USB converters with a drop of data during transfer to the computer. Additionally they do not amplify the low headset-microphone signal with sufficient quality. In the recording sets two indepent microphone amplifiers ART Tube MP are used. High level signals come to quasi-audiophile USB A/D converter M-Audio Transit and they are transferred to the computer through the USB interface. This two-channel configuration enables simultaneous recordings for 'close-' and 'middle distance'. In courtrooms, where the computer and its operator are not allowed to stay near the microphones, the wireless systems Sennheiser ew300G2 are used between microphones and audio interface.

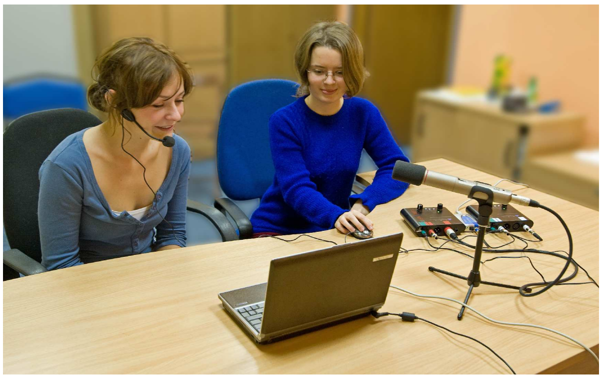
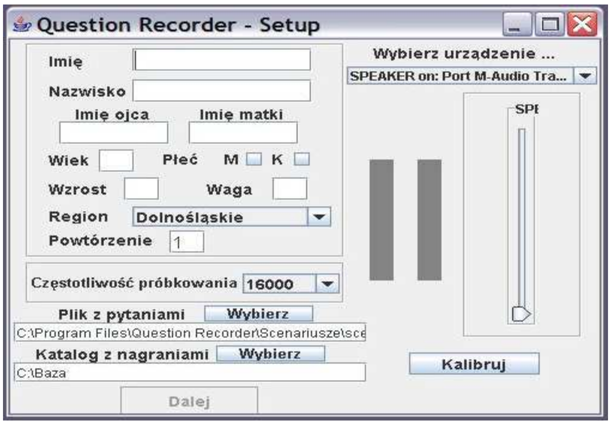


Figure 1. A recording session in an office

3.3 Software

To enable easy management of great number of speakers data and the recorded utterances the *QuestionRecorder* program was created using JAVA as the programming language. *QuestionRecorder* has two windows. The Setup Window (Figure 2) appears after program launches and requires setting of all necessary data concerning the recorded person (name, age, region of Poland, sex, weight, height), sampling rate (in fact fixed to 16kHz), ID number of the scenario (50 recording scenarios are available) and the directory for the recorded waveforms. The names of files are created automatically during the recording session. All the parameters are typically set only once at the beginning of each session. Before the beginning of the recording the audio track must be initially calibrated (recording level). With the Main Window (Figure 3) all (or only selected) utterances of a scenario may be recorded, with a possibility to check the recording quality or repeat them if needed. For each utterance two files are stored: a wave file and a text file describing recording conditions (SAM label file, cf. Fisher et al. 2000).

After finishing a series of recording sessions the speech data obtained from the *QuestionRecorder* software are stored on backup CDs and assessed (see p. 5.1 below) and then imported to *PPBW Annotation Database.*


Question Recorder - Setup

Imię

Nazwisko

Imię ojca

Imię matki

Wiek

Płeć M □ K □

Wzrost

Waga

Region: Dolnośląskie

Powtórzenie: 1

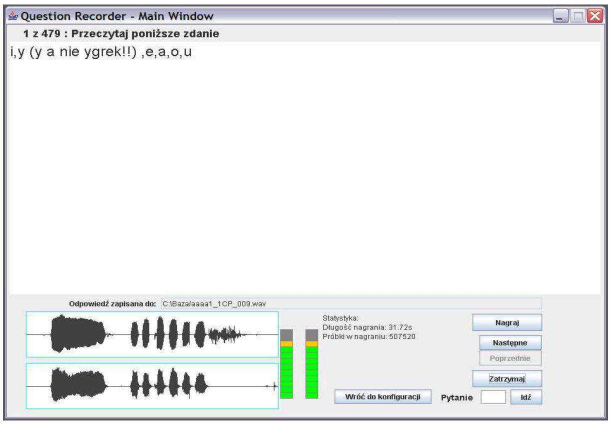
Częstotliwość próbkowania: 16000

Wybierz urządzenie ...: SPEAKER on: Port M-Audio Tra...

Plik z pytaniami: C:\Program Files\Question Recorder\Scenariusze\sce

Katalog z nagraniami: C:\Baza

Figure 2. *Question Recorder* – Setup Window


Question Recorder - Main Window

1 z 479 : Przeczytaj poniższe zdanie

i,y (y a nie ygrek!!) ,e,a,o,u

Odpowiedź zapisana do: C:\Baza/aaaa1_1CP_009.wav

Statystyka:
Długość nagrania: 31.72s
Próbki w nagraniu: 507520

Figure 3. *Question Recorder* - Main Window

4. Database Annotation

In the first stage the recordings are labeled by a group of 30 trained students of the Institute of Linguistics in Poznań whose work is supervised (and corrected if necessary) by a phonetician.

The second step is a thorough verification of the label files by a team of phoneticians accompanied by the automatic parsing of the files in order to synchronize the files contents with the lexicons.

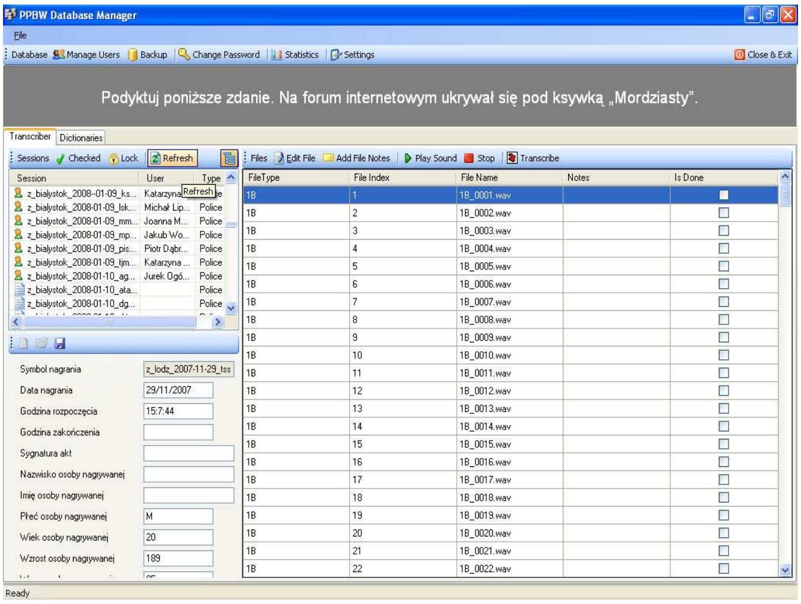
The lexicon created for the needs of the project consists of three parts: CW - common words, SAP - special application words and PN - proper names (Ziegenhain, et al., 2002). The CW lexicon (78.150 entries) covers a broad range of vocabulary extracted from an especially designed newspaper corpus (177.64634 words). For the SAP lexicon (5177 entries) we used various text sources: thematic dictionaries, technical documents and web portals to obtain vocabulary representative for a number of thematic areas. The PN lexicon consists of 46200 first/last names, organization and place names. Moreover, a frequency lexicon (Google-based word frequencies, 450.000 words) was designed to complete the coverage of the vocabulary occurring in the speech corpora.

After completion of the annotation verification, the quality of each utterance will be independently assessed based on a post-hoc automatic parsing (see the Prevalidation section below for more details).

Until now, 637 recording files have been included in the PPBW Annotation Database, 518 of them have been already annotated,and 140 of the annotated files are in mono (in the test phase), the remaining 378 files are in stereo.

4.2 Annotation Tools

For the purpose of the annotation of the recorded speech data new software was designed based on the Client-Server architecture using MSDE 2000, and Windows 2003 Server Client applications were programmed in C#. The tool was called *PPBW Annotation Database Manager* (cf. Figure 4) and is in charge of all the stages of the annotation procedure connected with sound and label files, text files, speaker information, lexicons search, and multi-user management. The program enables the import of the recordings produced with *QuestionRecorder* and the respective text files to the Annotation Database (after the database annotation is completed it will be possible to export all the files again to the required final format).The annotation solution is based on idea of only one working copy of data held on the server and client computers working as terminals. When the labelers log in to the server via *PPBW Annotation Database Manager* to work on a file, the file is downloaded from the server only for the edition time and then committed back to the server. All data exchange operations between client computers and the server are done automatically without using any additional storage devices. For the purpose of segmenting and labeling speech an open-source tool, *Transcriber,* version 1.5, was



Podyktuj poniższe zdanie. Na forum internetowym ukrywał się pod ksywką „Mordziasty”.

Session	User	Type
z_bialystok_2008--01-09_ks...	Katarzyna	[illegible]
z_bialystok_2008-01-09_lsk...	Michał Lip...	Police
z_bialystok_2008-01-09_mm...	Joanna M...	Police
z_bialystok_2008-01-09_mp...	Jakub Wo...	Police
z_bialystok_2008-01-09_pis...	Piotr Dąbr...	Police
z_bialystok_2008-01-09_tjm...	Katarzyna ...	Police
z_bialystok_2008-01-10_ag...	Jurek Ogó...	Police
z_bialystok_2008-01-10_ata...		Police
z_bialystok_2008-01-10_dg...		Police

Symbol nagrania	z_lodz_2007-11-29_tss
Data nagrania	29/11/2007
Godzina rozpoczęcia	15:7:44
Godzina zakończenia	
Sygnatura akt	
Nazwisko osoby nagrywanej	
Imię osoby nagrywanej	
Płeć osoby nagrywanej	M
Wiek osoby nagrywanej	20
Wzrost osoby nagrywanej	189

FileType	File Index	File Name	Notes	Is Done
1B	1	1B_0001.wav		□
1B	2	1B_0002.wav		□
1B	3	1B_0003.wav		□
1B	4	1B_0004.wav		□
1B	5	1B_0005.wav		□
1B	6	1B_0006.wav		□
1B	7	1B_0007.wav		□
1B	8	1B_0008.wav		□
1B	9	1B_0009.wav		□
1B	10	1B_0010.wav		□
1B	11	1B_0011.wav		□
1B	12	1B_0012.wav		□
1B	13	1B_0013.wav		□
1B	14	1B_0014.wav		□
1B	15	1B_0015.wav		□
1B	16	1B_0016.wav		□
1B	17	1B_0017.wav		□
1B	18	1B_0018.wav		□
1B	19	1B_0019.wav		□
1B	20	1B_0020.wav		□
1B	21	1B_0021.wav		□
1B	22	1B_0022.wav		□

Figure 4. *PPBW Database Manager* window

integrated in the system.

The database manager provides records of the working time with one-second-accuracy and enables generating working time statistics over a selected period of time.

Due to the confidential character of a part of the data the files are isolated from the Internet and protected from being copied from the system by unauthorized users. The central database is encoded and protected with a password. Annotation client computers are connected together in a private network. The labeler use ordinary user's accounts that do not allow for any configuration changes. Each of the labelers can access only the files processed by her or himself (authorized users can access and manage all recording data and user accounts).

Backup copies are created weekly and kept on separate hard disks which ensures the continuity of the work on annotation even in case of the server hard disk failure. Data are copied in a format enabling quick information retrieval at any time.

4.3 Annotation Specification

Annotation specification is based on SPEECON Deliverable 214 (Fisher et al., 2000).

Orthographic, case sensitive transcription is used in label files. Proper names are written with a capital letter. The proper names composed of several words are written with an underscore (e.g. Bielsko_Biała). White space is used as the word boundary markers. Phrase boundaries are not labeled by any special markers unless they coincide with pauses. Time section boundaries in the transcription files correspond to boundaries of continuous stretches of speech. For pauses longer than half a second the section boundary is obligatory.

Digit sequences are spelled out, with the exception of numbers being a part of certain proper names or application words which are labeled according to the lexicon. Letter sequences are in upper case, separated by a space. For letters realized by producing their phonetic form, slashes are used: /B/ /C/ ... /Z/. Polish digraphs are written with (only) the first letter capitalized. (e.g. Sz Cz or /Sz/ /Cz/ depending on realization). The letter Y is written: Y when pronounced /igrek/ or /ygrek/ and as /Y/ when pronounced /y/. For the transcription of e-mail and web addresses the lexicon is allowed to contain entries which are not meaningful words. The inflectional endings added to abbreviations, acronyms, application words or foreign names in Polish, are reflected in the label files (e.g. *Zapomniałem PINu* lit. *I forgot my PIN*). Foreign words are orthographically transcribed in their original spelling.

No punctuation is provided in the transcription other than the symbols used for special transcription purposes (Punctuation marks may occur in abbreviated names or application words like: CD-ROM or spółka z_o.o.). The punctuation provided to the speaker in the prompting text is held in the Annotation Database (together with the whole prompt text), however it is not inserted directly in the label files.

Words produced with extra or omitted syllables that are nevertheless intelligible are marked with one asterisk attached to the left of the mispronounced word (e.g. *pomyłka). The asterisk is not used for transcription of words representing careless pronunciation or normal dialectal or stylistic variation. Pronunciation variants will be covered in the lexicon partly based on the annotation files. Words, word fragments or other stretches of speech that are entirely unintelligible are transcribed as a sequence of two asterisks: "**" separated from neighbouring words with spaces.

Non-speech acoustic events are divided into four categories and transcribed as: filled pause, speaker noise,stationary noise or intermittent noise. Events are only transcribed if they are clearly distinguishable.The target speech signal is transcribed once for both left and right stereo channels as it is assumed that it remains the same for both of the channels (the possible delay of the speech signal between stereo channel is expected to be very small, i.e. 3 ms at most). The most important differences between channels come from noises, and are reflected in the transcription by indexes informing in which of the channel(s) a noise occurred (for example: [fil] - a filled pause observed in both channels, [fil:1] - a filled pause in the left channel, [fil:2] - a filled pause in the right channel). The insertion of the noise markers is semi-automatic (keyboard shortcuts are implemented in Transcriber).

Labelers may add comments on speaker characteristics or other features that are not included in the annotation specification, this information is stored in one of *PPBW Annotation Database Manager*'s fields.

5. Prevalidation

5.1 Recording quality assessment

The recordings are assessed by an expert phonetician with the help of a special tool: “Recording Checker” designed for the recording control procedure in the present project (cf. Figure 5). The most important characteristics of the program are as follows: a comfortable interface for listening to the recordings; easy navigation between recording sessions; volume measure module and distortion detector; session completeness control module; subjective assessment module (reading style, pronunciation, possible noises, reverberation, wrong microphone setup); session(s) assessment reports.

5.2 Annotation verification and dictionary supplement

Files annotated by students have been searched for tokens that are not included in the project's lexicons. The resulting word list is checked manually by an expert and the tokens will be either corrected in the label files oradded to the lexicons.

All label files produced by students are inspected by a team of phoneticians following the same guidelines as the student labelers. At this stage two more attributes are added to the recording file information held in an additional field of the PPBW Annotation Database: the subjective assessment of the speech rate (too fast or too slow) and the speech quality (very careless or non-standard pronunciation or speech disorders); these attributes are to be assigned only when in the expert's opinion the recording deviates to a great extent from the norm.

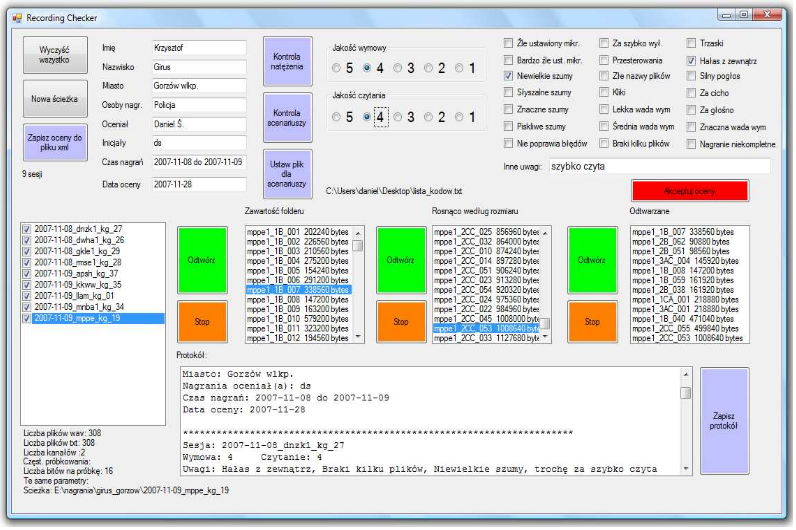
Finally, the quality of each recording will be assessed independently, based on the final parsing of the annotation files. According to SPEECON deliverable D214 (Fischer et al. 2000), each recording file will obtain one of four grades (*garbage, noise, other, OK*) depending on the amount and type of noise markers included in its corresponding label file.

6. Future work

It will be possible to provide the general statistics for the database after the annotation of the variable part of the database. The evaluation process by an independent centre (e.g. ELRA) should estimate the quality and usefulness of the database for building ASR system for Polish.

7. Conclusions

The JURISDIC speech dictation database is designed to provide material for both training and testing of speech dictation of common and legal texts which include



Imię: Krzysztof
Nazwisko: Girus
Miasto: Gorzów wlkp.
Osoby nagr.: Policja
Oceniał: Daniel Ś.
Inicjały: ds
Czas nagrań: 2007-11-08 do 2007-11-09
Data oceny: 2007-11-28

9 sesji

Jakość wymowy: ○ 5 ● 4 ○ 3 ○ 2 ○ 1
Jakość czytania: ○ 5 ● 4 ○ 3 ○ 2 ○ 1

□ Źle ustawiony mikr. □ Za szybko wył. □ Trzaski
□ Bardzo źle ust. mikr. □ Przesterowania ☑ Hałas z zewnątrz
☑ Niewielkie szumy □ Złe nazwy plików □ Silny pogłos
□ Słyszalne szumy □ Kliki □ Za cicho
□ Znaczne szumy □ Lekka wada wym □ Za głośno
□ Piskliwe szumy □ Średnia wada wym □ Znaczna wada wym
□ Nie poprawia błędów □ Braki kilku plików □ Nagranie niekompletne

Inne uwagi: szybko czyta

C:\Users\daniel\Desktop\lista_kodow.txt

☑ 2007-11-08_dnzk1_kg_27
☑ 2007-11-08_dwha1_kg_26
☑ 2007-11-08_gkle1_kg_29
☑ 2007-11-08_rnse1_kg_28
☑ 2007-11-09_apsh_kg_37
☑ 2007-11-09_kkww_kg_35
☑ 2007-11-09_llam_kg_01
☑ 2007-11-09_mnba1_kg_34
☑ 2007-11-09_mppe_kg_19

Zawartość folderu

```
mppe1_1B_001  202240 bytes
mppe1_1B_002  226560 bytes
mppe1_1B_003  210560 bytes
mppe1_1B_004  275200 bytes
mppe1_1B_005  154240 bytes
mppe1_1B_006  291200 bytes
mppe1_1B_007  338560 bytes
mppe1_1B_008  147200 bytes
mppe1_1B_009  163200 bytes
mppe1_1B_010  579200 bytes
mppe1_1B_011  323200 bytes
mppe1_1B_012  194560 bytes
```

Rosnąco według rozmiaru

```
mppe1_2CC_025  856960 bytes
mppe1_2CC_032  864000 bytes
mppe1_2CC_010  874240 bytes
mppe1_2CC_014  897280 bytes
mppe1_2CC_051  906240 bytes
mppe1_2CC_023  913280 bytes
mppe1_2CC_054  920320 bytes
mppe1_2CC_024  975360 bytes
mppe1_2CC_022  984960 bytes
mppe1_2CC_045  1008000 byte
mppe1_2CC_053  1008640 byte
mppe1_2CC_033  1127680 byte
```

Odtwarzane

```
mppe1_1B_007  338560 bytes
mppe1_2B_062  90880 bytes
mppe1_2B_051  98560 bytes
mppe1_3AC_004  145920 bytes
mppe1_1B_008  147200 bytes
mppe1_1B_059  161920 bytes
mppe1_2B_038  161920 bytes
mppe1_1CA_001  218880 bytes
mppe1_3AC_001  218880 bytes
mppe1_1B_040  471040 bytes
mppe1_2CC_055  499840 bytes
mppe1_2CC_053  1008640 bytes
```

Protokół:

```
Miasto: Gorzów wlkp.
Nagrania oceniał(a): ds
Czas nagrań: 2007-11-08 do 2007-11-09
Data oceny: 2007-11-28

***************************************************************************
Sesja: 2007-11-08_dnzk1_kg_27
Wymowa: 4     Czytanie: 4
Uwagi: Hałas z zewnątrz, Braki kilku plików, Niewielkie szumy, trochę za szybko czyta
```

Liczba plików wav: 308
Liczba plików txt: 308
Liczba kanałów :2
Częst. próbkowania:
Liczba bitów na próbkę: 16
Te same parametry:
Scieżka: E:\nagrania\girus_gorzow\2007-11-09_mppe_kg_19

Figure 5. *Recording Checker* interface

isolated word systems, wordspotting systems and vocabulary independent systems which use either whole-word or sub-word modeling approaches. This, together with the substantial size of the speech corpus is expected to provide sufficient research material for LVCSR development.

7. Acknowledgements

The project is supported by The Polish Scientific Committee (Project ID: R00 035 02).

8. References

Demenko G., Wypych M., and Baranowska, E. (2003). Implementation of Grapheme-to-Phoneme Rules and Extended SAMPA Alphabet in Polish Text-to-Speech Synthesis. Speech and Language Technology, Edition PTFON, vol.7.

Djamel Mostefa, Olivier Hamon†, Khalid Choukri,Van den Heuvel, H., Choukri, K., Gollan, Chr., Moreno,A., Mostefa, D. (2006). TC-STAR: New language resources for ASR and SLT purposes. In: Proceedings of the LREC 2006, Genoa, Italy.

Docio-Fernandez Laura, Antonio Cardenal-Lopez, Carmen Garcia-Mateo, TC-STAR 2006 Automatic Speech Recognition Evaluation: The UVIGO System, TC_STAR Workshop on Speech-to-Speech Translation, June 19–21, 2006, Barcelona.

ELRA: European Language Resources Association homepage: http://www.elra.info/

Fischer, V., Diehl, F., Kiessling, A., Marasek, K. 2000. Specification of Databases - Specification of annotation. SPEECON Deliverale D214.

Gibbon D., Moore R., Winski R., Handbook of Standards and Resources for Spoken Language Systems, deGruyter, 1997.

Henk van den Heuvel, Eric Sanders, Validation of language resources in TC-STAR, TC-STAR Workshop on Speech-to-Speech Translation, Barcelona, June 19–21, 2006.

JURISDIC project and Laboratory of Speech and Language Technology website: http://www.speechlabs.pl

Loof J., Ch. Gollan, S. Hahn, G. Heigold, B. Hoffmeister, Ch. Plahl, D. Rybach R. Schluter and H. Ney, The RWTH 2007 TC-STAR Evaluation System for European English and Spanish, Interspech 2007, 2145-1249.

SPEECON: http://www.speechdat.org/speecon/index.html

TRANSCRIBER: http://trans.sourceforge.net/

Sundermann D. A language Resources Generation Toolbox for Speech Synthesis, TC_STAR publication, http://www.tc-star.org/pubblicazioni/scientific_ publications/Siemens/2005/ast2005.pdf

TC STAR project homepage: http://www.tc-star.org /

Ziegenhain, U. et al. 2002. Specification of corpora and word lists in 12 languages. LC-STAR Deliverable D1.1.