The EURONOUNCE corpus of non-native Polish for ASR-based Pronunciation Tutoring System

Natalia Cylwik, Agnieszka Wagner, Grażyna Demenko

Adam Mickiewicz University, Institute of Linguistics, Department of Phonetics,
Poznań, Poland
{nataliac, wagner, lin}@amu.edu.pl

Abstract

This paper gives a detailed information on the design of the speech corpus for the purpose of developing an ASR-based pronunciation tutoring system. In the first place, assumptions on the structure of the corpus are presented. Then collection of text material, recordings and procedure of annotation of the resulting speech corpus are described. In the end, preliminary results of the analysis of pronunciation errors are discussed. They provide information which is important for ASR training and testing on the one hand, and automatic error detection on the other hand.

1. Introduction

Major advance in the area of Speech and Language Technology which we have witnessed in recent years has offered new potential uses to the field of second language acquisition, which coincided with a growing interest in L2 pronunciation and prosody, to this point largely neglected in foreign language teaching. These changes resulted in the development of Computer-Assisted Pronunciation Training (CAPT) systems using modern techniques such as automatic speech recognition and automatic error detection. Although yet a few years ago it was often doubted whether ASR-based systems were reliable enough to recognize non-native speech [1], advanced CAPT systems such as ISLE [2], PLASER [3], SRI EduSpeakTM [4] and AzAR [5] report on a recognition accuracy comparable to that of native speech and a fairly satisfactory detection of pronunciation errors [6], [3], which can be largely credited to the fact that in innovative systems ASR technology is trained on and tested against non-native speech.

Since most L2 pronunciation errors result from interference with the learner's L1 [7], [8] the L1 of the potential end-users of the software-to-be needs to be specified beforehand, which should be followed by the creation and annotation of a database containing L1, L2 and non-native speech. Although the number of CAPT systems as well as the interest in creating large speech corpora are growing, the literature on the design and annotation of non-native speech corpora is still scarce. The article provides a detailed description of the design and annotation of a speech corpus for the pair L1 German, L2 Polish (DE-PL), collected as part of a larger multilingual speech database created within the scope of the Euronounce project (http://www.euronounce.net). The project aims at creating an Intelligent Language Tutoring System with multimodal feedback functions, which will add a fresh approach to pronunciation learning consisting in offering learners multimodal feedback based on speech recognition and automatic error detection, user-friendly design and interactive exercises including prosody. The project focuses on Polish, Russian, Czech and Slovak as L2 for German learners and on German as L2 for learners with one of the above mentioned L1.

2. The corpus

Non-native speech corpora including target languages other than English are still very sparse and those that exist are usually very limited in size and hardly accessible, created mostly for the purpose of analyzing phonetic interferences. In this context Euronounce corpus is unique as it includes Polish as a target language and has been created for purposes of an ASR-based CAPT system, which requires considerable size and complexity.

Most non-native speech corpora have been created for either of the three purposes: to study phonetic interference [10], to study L2 acquisition processes [8] or to create ASR-based CAPT systems [2], [3], [4], [5]. It seems crucial to precisely determine the purpose of the corpus since it influences its design. In the first two cases, the design of the L2 corpus can remain L1-independent, merely representing L2 phonetic system, i.e. it can be void of any premises. The development of an ASR-based CAPT system requires building complex language corpora that would contain phonetically rich and balanced sentences in both L1 and L2 for ASR training and testing on the one hand, and provide enough evidence of pronunciation errors typical of speakers with a particular L1 for purposes of automatic error detection and ASR on the other hand. The purpose is not, therefore, to find out what errors are characteristic of L2 learners with a particular L1 but to collect most common errors and systematic errors, typical of the language pair under consideration.

Following these assumptions, in the development of the Euronounce speech database for the pair DE-PL several sub-databases were created, whose exact structure and purpose are described in the next section.

2.1. Non-native speech database

It serves collecting evidence of most common pronunciation and prosodic errors made by German learners of Polish.

2.1.1. The procedure of text material collection

In the development of the text material for the database special emphasis was put on using simple vocabulary and grammatical structures adjusted to elementary students. This approach was based on the assumption that speakers should understand what they read and a potential source of errors should be the phonetic structure of the text material rather

than the lexical, syntactic or semantic one. Needless to say, this approach constituted a considerable constraint on the selection of the material for the recordings. In some cases it resulted impossible to meet the requirement of lexical and syntactic simplicity since it was considered of greater importance to elicit certain phenomena. Therefore, part of the corpus was addressed to upper-intermediate and advanced students only.

2.1.2. The structure of non-native speech database

The non-native corpus is comprised of 6 tests:

Accent test – crucial part of the non-native corpus since it takes into account the speakers' L1, being a collection of sentences containing those Polish sounds and phonetic phenomena that are considered difficult from the point of view of a German learner, e.g. Polish [x] in words such as 'ich' (Eng. 'their') which Germans might pronounce as [C]. It was created by teachers of Polish as a second language on the basis of their practical experience in teaching Polish to Germans and a comparison of German and Polish phonetic systems. The second method was based on the assumption that most pronunciation errors are systematic and possible to be predicted by looking into phonetic systems of L1 and L2 [2], [7]. The test contains 125 sentences.

Dialectological test – 124 sentences containing words with alternative pronunciations according to the dialect spoken, e.g. bank pronounced as /bank/ or /baNk/, as well as representing Polish assimilation processes within words and at word boundaries, e.g. bluzka (Eng. 'blouse') pronounced as /bluska/ and a full range of Polish phonemes in different contexts, word and sentence positions, vowels in minimal pairs, e.g. tik:tak (Eng. 'tick':yes'), consonants in oppositions voiceless vs. voiced, e.g. pić:bić (Eng. drink:hit), vowels in stressed and/vs. unstressed positions e.g. ma:mama (Eng. 'has':,mum'), etc. This part is independent of L1, which enables detection of unpredicted and possibly L1-unrelated errors. Although all mispronunciations, including the uncommon ones, are annotated, only frequent errors will be taken into account while training and testing ASR and automatic error detection since in order to be reliable the system needs to be trained on numerous occurrences of the same error, possibly coming from different speakers.

Spontaneous speech test – addressed only to more advanced students. It consists of four simple tasks such as finishing a sentence, e.g. 'My hobby is…' and explaining the meaning of a proverb or idiomatic expression commonly known both in Poland and Germany, e.g. Pol. 'przemoknąć do suchej nitki', Germ. 'keinen trockenen Faden (mehr) am Leibe haben' (Eng. 'to get soaked to the skin'). The primary goal of this test is to assess speakers' proficiency level and to investigate phenomena characteristic of this mode of speech. Although spontaneous speech is found to reveal different pronunciation errors than read speech [8], the current stage of non-native speech recognition development does not allow a reliable recognition of spontaneous speech by non-native speakers. Therefore, this part of the corpus will not be subjected to a detailed analysis of segmental errors.

Continuous speech test – three passages (72 sentences altogether) taken from stories by H. Ch. Andersen and Grimm Brothers, two of which are addressed to upper-intermediate and advanced students only. The aim is to collect evidence of prosodic and discourse-level errors. All the texts are phonetically rich and balanced and serve also ASR training purposes.

Prosody test – a set of 59 sentences which aim at collecting evidence of those prosodic errors that are most easily detectable and most crucial for comprehension, such as erroneous stress placement or non-native-like vowel duration. Other prosodic features included in the test are: intonation in neutral sentences vs. sentences with focus, in questions vs. statements, commands and requests, etc.

Phondat corpus - it contains three sets of phonetically rich and balanced sentences (341 altogether) for the purposes of ASR training and testing and collecting mispronunciations of consonant clusters. Polish is an exceptionally consonantal language allowing for sequences of even four or five consonants in a word, e.g. drgnąć (Eng. 'to quiver'), which is often a source of pronunciation errors in foreigners. The three sets were diversified as regards the level of difficulty of the vocabulary and grammatical structures used. Only one set was addressed to elementary, pre-intermediate and intermediate students.

2.1.3. Statistical analysis of the text corpus

The non-native corpus includes 721 sentences to read by 18 German learners of Polish. An overview of word and triphone coverage for different parts of the German-Polish corpus is given in the Fig. 1 below.

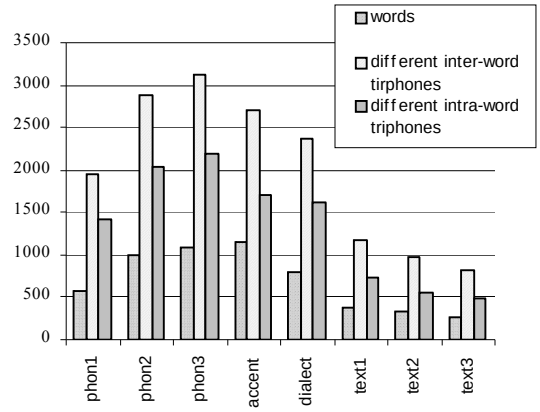

Figure 1: Triphone and word coverage in the text corpus.

2.2. Source-language database

It serves as a reference for the assessment of non-native pronunciation. It contains Polish speech by 18 Polish native speakers reading part of the text material intended for non-native corpus, namely Phondat test and Prosody test.

2.3. Training database

50 hours of Polish and 50 hours of German speech provided by over 100 Polish and German native speakers for the general speech recognizer training.

2.4. Reference database

It will contain Polish speech by one male and one female speaker with pleasant voices. The text material will include isolated words, minimal pairs, sentences, mini dialogues and

continuous speech passages which will be implemented into the tutoring system as curriculum for German learners of Polish.

3. Recording procedure

Speakers for the non-native speech database were mostly German students studying temporarily in Poland. All speakers were asked to fill in a questionnaire which was to provide standard personal data as well as information on their experience related to Polish and their writing, speaking, reading, grammar and pronunciation skills. On this basis speakers were assigned to different proficiency groups from A1-C2, according to Common European Framework of Reference for Languages. Speakers with the level A1-B1 were recorded during 2 sessions whereas more advanced students (level B2-C2) needed 4 sessions. Altogether 18 speakers were recorded with a balanced distribution of proficiency level and gender i.e., 6 speakers (3 males and 3 females) per level A, B and C. Speakers were provided with the material beforehand (except for spontaneous speech test) and at the same time they were discouraged from practicing pronunciation. The speakers could stop the recording at any time during the session and re-record sentences which they themselves considered mispronounced. As the primary goal was to collect only pronunciation errors resulting from inability to pronounce certain words or sounds or from unawareness that they should be pronounced in a different way, whenever mistakes resulting from slipping of the tongue, misreading or too long hesitation pauses occurred, the sentence was re-recorded.

4. Annotation

4.1. Annotation procedure

The annotation contains transcription of the speakers' actual utterances in relation to a reference transcription containing canonical native pronunciation which was generated automatically and then manually verified. The annotation of the DE-PL speech corpus has been performed by four labelers - three native speakers of Polish (phoneticians) and one native speaker of German (expert in Slavic languages).

The annotation procedure is similar to that proposed in [11] and involves the following steps. Firstly, a trained phonetician – a native speaker of the target language (here: Polish) verifies and corrects the canonical transcription which is then used to produce a phonetic segmentation. Subsequently, the annotator identifies the portions of the signal perceived as mispronounced and marks deviations from the canonical pronunciation. At the phone level three kinds of pronunciation errors are distinguished: substitutions, insertions and deletions.

The resulting annotation is verified by a native speaker of the source language (here: German) who confirms or rejects the results of assessment of pronunciation errors provided by other annotators.

Each time a deviation from the canonical pronunciation is observed the annotator can choose from among two phoneme sets: a modified version of Polish SAMPA [12] and extended SAMPA for German [13]. A special set of labels is provided to describe approximations to Polish or German phonemes and the diacritics available in the IPA alphabet are used to describe specific phonetic and articulatory phenomena.

4.2. Statistical and linguistic analysis of non-native speech database – preliminary results

The statistics computed on the basis of 880 fully annotated utterances (62 minutes) provided by 6 students of Polish with L1 German (4 advanced and 2 beginners) have shown that most pronunciation errors involve substitutions (3446 instances) whereas deletions and insertions are less frequent (934 and 842 respectively). Unexpectedly, no correlation between students' proficiency level and the number of pronunciation errors made was observed. The figure below shows the distribution of different types of errors in two different proficiency groups normalized with respect to the number of students in each group (in percentages); values on the plot give the absolute number of errors in each group.

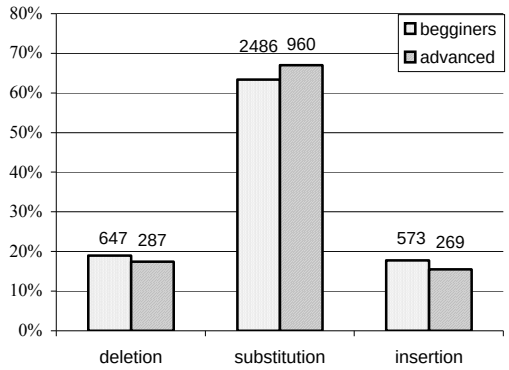

Figure 1: Frequency and distribution of pronunciation errors in different proficiency groups.

4.2.1. Substitutions

The analysis of the types of substitutions revealed that German learners of Polish tend to reduce final unstressed vowels and have problems with realization of fricatives and affricates not present in German. They also substitute difficult consonantal clusters such as plosive + fricative for affricates e.g. /tSeba/ (one must) pronounced as /t^Seba/.

Most speakers realize voiced fricatives, affricates and plosives similarly to German lenes consonants (perceptually salient partial or total devoicing). Other common substitutions found in the corpus are shown in Tab. 1.

The substitutions discussed here are easily predictable as they mostly result from interferences between students' L1 and L2 or transfers of pronunciation regularities from German to Polish.

4.2.2. Deletions

The most common type of deletion occurred in consonant sequences which were difficult to pronounce. Overview of the most common deletions found the part of the DE-PL speech corpus analyzed so far is given in Tab. 2.

Table 1: Common substitutions found in the DE-PL corpus.

text	Canonical pronunciation	Substi-tution	Example
ą	/o/+{/m/, /n/, /n'/, /N/, /w~/, /j~/}	O~	/mow~S/ (husband)
ę	/e/+{/m/, /n/, /n'/, /N/, /w~/, /j~/}	E~	/mew~sci/ (male)
ał	/aw/	/aU/	/mjawa/ (she had)
oj	/oj/	/OY/	/ojt^s'et^s/ (father)
aj	/aj/	/aI/	/jajko/ (egg)
r	/r/	/6/	/varSava/ (Warsaw)

Table 2: Common deletions (- marks the deleted segment).

Deleted segment	Previous context	Example
plosive	plosive, fricative, affricate	/paj~st-fo/ (state)
fricative	affricate, fricative	/bezv-zglend-ny/ (absolute)
nasal	nasal	/obron-ny/ (defensive)
/r/ and /l/	various	/tyl-ko/ (only)

4.2.3. Insertions

More than 43% of insertions occurred as a result of realizing a vocal onset with a glottal stop which is untypical of Polish where soft vocal onset occurs most of the time. Consequently, the most frequent insertion was that of a glottal stop between vowels at syllable or morpheme boundaries e.g. /ot^se-Qan/ (Eng. 'ocean'; - marks the inserted segment) and at the beginning of words starting with a vowel e.g. /awto/ (Eng. 'car', pronounced as /-QaUto/.

One fifth of all insertions involved pronounciation of vowels inside consonant clusters which speakers found difficult to pronounce e.g. /f-yt^Soraj/ (Eng. 'yesterday').

In open syllables starting with palatalized consonants some of the less advanced speakers tended to de-palatalize the consonant and to insert /j/ after it (18% of all insertions) e.g. /f-y_pecin'-n-je/ (Eng. 'in Peking'; the first and third hyphen indicate insertions of /y/ and /j/ after /f/ and /n/ respectively, whereas the second one - substitution of /n/ for palatalized /n'/).

5. Conclusions

Creation of a non-native corpus is the first and necessary stage in the development of CAPT systems as important as the other stages which include the development of ASR technology, design of interface and creation of a curriculum.

This paper provided a detailed description of the structure, content, creation and annotation of the Polish non-native speech database designed for the purposes of Euronounce Intelligent Language Tutoring System. The database provides evidence of the most common pronunciation errors made by German learners of Polish, which will serve as a basis for ASR training and testing and design of curriculum aimed specifically at German students.

6. Acknowledgements

This project has been funded with support from the European Commission within the Lifelong Learning Programme (project 135379-LLP-1-2007-1-DE-KA2-KA2MP). The project homepage is located at: http://www.euronounce.net.

7. References

[1] Liu, M., Moore, Z., Graham, L. and Lee, S., "A Look at the Research on Computer-Based Technology Use in Second Language Learning: A Review of the Literature from 1990–2000". *Journal of Research on Technology in Education*, 34(3):250-273, 2002.

[2] Atwell, E., Howarth P. and Souter, C., "The ISLE Corpus: Italian and German Spoken Learners' English", *ICAME Journal*, 17:5-18, 2003.

[3] Mak, B., Siu, M., Ng, M., Tam, Y., Chan, Y., Chan, K., Leung, K., Ho, S., Wong, J. and Lo, J., "PLASER: Pronunciation Learning via Automatic Speech Recognition". *Proc. HLT-NAACL Workshop on Building Educational Applications using Natural Language Processing 2003*.

[4] Franco, H. Abrash, V. Precoda, K. Bratt, H. Rao, R., Butzberger, J., Rossier, R. and Cesari, F., "The SRI EduSpeak(TM) System: Recognition and Pronunciation Scoring for Language Learning". *Proc. InSTIL 2000* Dundee, Scotland, pp. 123-128.

[5] Jokisch, O., Koloska, U., Hirschfeld, D. and Hoffmann, R. "Pronunciation learning and foreign accent reduction by an audiovisual feedback system". *Proc. 1st Intern. Conf. on Affective Computing and Intelligent Interaction (ACII),* Beijing, China, 2005, pp. 419-425.

[6] Neri, A., Cucchiarini, C. and Strick, W., "Automatic Speech Recognition for second language learning: How and why it actually works". *Proc.15th ICPhS*, Barcelona 2003.

[7] Wells, J.C., "Overcoming phonetic interference. English Phonetics", *Journal of the English Phonetic Society of Japan,* 3:9-21, 2000.

[8] Flege, J.E., Munro, M.J. and MacKay, I.R.A., "Factors affecting degree of perceived foreign accent in a second language". *J. Acoust. Soc. Am.* 97:3125-3134, 1995.

[9] Cylwik, N., Demenko, G., Jokisch, O., Jäckel, R., Rusko, M., Hoffmann, R., Ronzhin, A., Hirschfeld, D., Koloska, U. and Hanisch, L., "The use of CALL in acquiring foreign language pronunciation and prosody – general specifications for Euronounce Project". *Proc. SASR*, Piechowice, Poland, September 2008.

[10] Szalkowska, E., "Epenthesis as Korean learners' strategy in the acquisition of Polish consonant clusters", *Studia Phonetica Posnaniensia*, 8: 41-71, 2007.

[11] Bonaventura, P., Herron, P. and Menzel, W., "Phonetic rules for diagnosis of pronunciation errors". *Proc. KONVENS 2000*, Ilmenau, October 2000.

[12] Demenko G., Wypych M. and Baranowska E., "Implementation of Polish grapheme-to-phoneme rules and extended SAMPA in Polish TTS synthesis". *Speech & Language Technology*, 7:79-96, 2003.

[13] Bavarian Archive for Speech Signals: Extended German SAMPA, Retrieved on 11th November 2008 from: http://www.phonetik.uni-muenchen.de/forschung/Bas/BasSAMPA