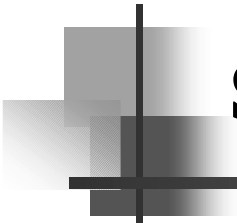


PELT: An English language tutorial system for Polish speakers



*Katarzyna Dziubalska-Kołodziej,
Anna Bogacka, Dawid Pietrala,
Mikołaj Wypych, Grzegorz Krynicki*



Institutions

- Adam Mickiewicz University, Poznań, Poland
 - School of English
 - Center for Speech and Language Processing
- Polish Academy of Sciences
- AMU grant

cslp@ifa.amu.edu.pl



PELT

- The Polish-English Literacy Tutor
- a multimodal multilingual tutorial system for foreign language learning
- English for adult Polish learners
- requires a specific speech recognition system dealing with highly accented, strongly variable second language speech



Plan of the talk

1. Introduction
2. Phonetic characteristics of Polish English
3. Corpus preparation & annotation
4. Corpus statistics
5. Automatic error detector



Introduction



CLT

- Colorado Literacy Tutor: the platform for developing PELT
- Center for Spoken Language Research (CSLR), University of Colorado, Boulder
- automatic speech recognition (SONIC), dialog systems and animated agents



PLT

- Polish Literacy Tutor - a tutorial system for native Polish (a prototype)
- the corpus for SONIC: 113 speakers
- Polish visemes (lip shapes for phonemes) of three speakers were video-recorded and matched with the English visemes
- SAMPA mappings of the English phonemes to visemes were adapted and used for phonetically similar Polish phonemes



PELT

- highly accented speech
- variable speech depending on the level of proficiency
- difficulty in aligning acoustic features to phonemes

Phonetic characteristics of Polish English



Polish vs. English phonology: typology

- Polish is not stress-timed
 - vowels – tend to maintain their quality
 - word stress – in Polish it is fixed on a penultimate syllable
 - consonantal clusters – Polish is much richer in clusters in all word positions
- Polish is not weight-sensitive
- the segmental inventory of Polish is much nearer to the average balance between vowels and consonants (ca. 6 to over 20)



Polish vs. English phonology: system adequacy

- the inventory of Polish vowels is entirely different
- in consonants, there are important systemic & distributional differences
 - Polish lacks dental apical fricatives while it has dental laminal obstruents
 - the distribution of a velar nasal is restricted to homorganic pre-velar-stop contexts



Polish vs. English phonology: universals

- Polish is unmarked with reference to the process of word final obstruent devoicing
- as well as interconsonantal voice agreement



Polglish pronunciation: predictions

- the errors will either be directly L1-induced (i.e. caused by the interference of the system-adequate features of Polish)
- or caused by the type-specific or universal processes



Illustration of errors

- L1-induced: a substitution of a Polish dental or labio-dental obstruent (fricative or stop) for the English apical dental fricative
- typological: the inability to reduce unstressed vowels, the difficulties in stress placement
- universal: word-final obstruent devoicing (a universal phonological process reinforced in Polish speakers by the system-adequacy)

Corpus preparation & annotation





Proficiency level

- the speakers will be divided into proficiency groups by means of statistical tests performed on the number and quality of errors they make
- the speech of any user beginning to use the program will be compared to the group characteristics and the users will thus receive training at the appropriate level



Prompts

- sentences which had been used for recording native American English speakers
- designed to ensure maximum diversity of phonetic contexts



Speakers

- 116 speakers (85 females and 31 males)
- age ranged from 16 to 43, with the mean age 21,9 years and standard deviation 4,4 years
- 24% - the First Certificate in English level, 62% - the Cambridge Advanced Certificate in English level, & 14% - the Cambridge Proficiency Examination level
- 71,6% declared to have been learning British English accent, 27,6% American English accent, and 0,9% were hesitant
- subjects were asked to name geographical regions they came from and other foreign languages they spoke



Annotation

- students of English who completed a two year course in English phonetics
- listen to the recordings, compare them to all its acceptable native readings and annotate the differences by means of a predefined tagging notation
- “all acceptable native readings”
 - all pronunciations accepted by educated native speakers of the standard variety of English identical to the variety declared by the subject, i.e. Received Pronunciation (RP) or General American (GA)
 - produced without disfluencies and noises




Corpus statistics



The analysis of transcripts

- the analysis of 100 transcripts (of the 116)
- departures from the transcript in the speech of the subjects
 - phonetic: 10 error types grouped into 7 major categories
 - non-phonetic: word-level errors, disfluencies, restarts and noises

Phonetic error type frequencies in PELT (1)

	Error type	Source of likely Polish error (error in brackets)	Count	%
CO NS ON AN TS		velar nasal (/ɲ/, /ɲ̥/, /ɲ̥/) e.g. everything */ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥/	360	5,0
		/ɲ̥/+V with no /ɲ̥/ (/ɲ̥ ɲ̥ V/) e.g. singer */ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥/		
	voicing of consonants and voicing of consonant clusters	voiced /ɲ̥ ɲ̥ ɲ̥/ or /ɲ̥ ɲ̥ ɲ̥/ (/ɲ̥/) e.g. this boy */ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥/	2429	33,4
		final voiced obstruent (devoicing) e.g. disguise */ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥/		
		voiced obstruent + /ɲ̥/ or /ɲ̥/ + voiced obstruent (regressive assimilation) e.g. absurd */ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥/		
	consonant clusters	/ɲ̥ ɲ̥ ɲ̥/, /ɲ̥ ɲ̥ ɲ̥/ etc. word-finally (schwa insertion) e.g. attached */ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥/	14	0,2
place of articulation	/ɲ̥/ → /ɲ̥/, /ɲ̥/ etc., /ɲ̥/ → /ɲ̥/ etc. (except /ɲ̥/ → /ɲ̥/) e.g. think */ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥/	812	11,2	
manner of articulation	/ɲ̥/ → /ɲ̥ ɲ̥/ etc. e.g. cliché */ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥ ɲ̥/	49	0,7	

Phonetic error type frequencies in PELT (2)

VO WE LS	/★/ or /ǂ□/	schwa quality and/or quantity e.g. cater */&ℳ 𐀃 𐀄 𐀅/	2316	31,8
	monophthongs	vowel quality error, vowel nasalisation e.g. fenced *[↗ ↘ ↙ ↚ ↛ ↜ ↝]		
	di/triphthongs	/ℳ ★/ or /𐀃 ★/ (/er/ breaking, schwa) in RP e.g. tier */𐀄 𐀅 𐀆★/		
		/† ★/ (/♦/ breaking, schwa, /◆/) in RP e.g. poor */□ † ‡ ★/		
OT HE R	word stress	stress placement errors e.g. astronomy */→ ◆ ◆ □ ★ ↗ ■ ① ○ 𐀃/	1287	17,7
		secondary stress (reduced to unstressed) e.g. impartiality */𐀃 ○ □ ★ 𐀄 ↗ → ● ★ ◆ 𐀃/		
	variety of English	inconsistence in the use of RP or GA e.g. after */↗ → ↘ ◆ ★/ instead of /↗ 𐀃 ↘ ◆ ★/ or /↗ → ↘ ◆ ★ □/		
total			7277	100



Word-level errors

- the total of **1478**
 - deletions – 23,2%
 - insertions – 23,4%
 - word order errors – 0,5%
 - misreadings – 33,4%
 - substitutions – 19,5%



Disfluencies, restarts & noises

- total of **491** disfluencies
 - 50,1% pauses
 - 39,8% hesitated chunks
 - 10,1% fillers
- **526** restarts
- **544** noises (aside remarks, audible inhaling or exhaling, laughter, cough, throatclear, sniffing, steps, etc.)



Automatic error detector (1)

- the presented speech corpus is used as training data for automatic pronunciation errors detector
- the goal of the detector is to automatically determine the type (and possibly intensity) of pronunciation errors occurring in English speech produced by Polish native speakers
- the phonetic error typology constitutes the basis for the preparation of accompanying acoustic models and pronunciation models



Automatic error detector (2)

- the detector, given an acoustic observation sequence and an orthographic transcript evaluates the observation sequence using each of the acoustic and pronunciation models
- the resulting scores for each model allow to measure the intensity of pronunciation error by comparing the score of the error model to the score of the native English model
- for the purpose of scoring comparable additional normalization factors need to be extracted from the acoustic and pronunciation models

CSLR characters



Ms. ReadWrite



Marni



Julie



Singo

Marge



Motion capture and synthesis





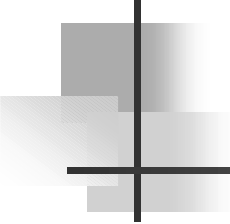
Lori Loudmouth

Lori Ramig



Animated Voice Therapist





Demo of PLT (Polish Literacy Tutor)

- potentially, already some PELT classification results, too