General and Task-Specific Corpus Resources for Polish Adult Learners of English

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

School of English & Center for Speech and Language Processing
Adam Mickiewicz University, 61-874 Poznań, al. Niepodległości 4
cslp@ifa.amu.edu.pl

Abstract

This paper offers a comparison of two resources for Polish adult learners of English. The first has been designed for Polish-English Literacy Tutor (PELT), a multimodal system for foreign language learning, as training input to speech recognition system for highly accented, strongly variable second language speech. The second corpus is a task-specific resource designed in the PELT framework to investigate the vowel space of English produced by Poles. Presented are linguistically and technologically challenging aspects of the two ventures and their complementary character.

1. Introduction

Development of a pronunciation tutorial system for Polish adult learners of English required annotated speech corpora of English produced by Polish speakers. Educational and technological requirements of second language teaching and learning were the motivation to create two new, high quality resources: the Polish English Literacy Tutor (PELT) corpus and the diphthong corpus, allowing for investigation of typological differences between phonological systems of English and Polish. The PELT corpus has been developed specifically to serve as a training input to the Polish English Literacy Tutor, which is multimodal and multilingual tutorial system for foreign language learning, requiring a specific speech recognition system dealing with highly accented, strongly variable second language speech. The PELT corpus is thus a general corpus designed to investigate all basic features of Polish English required by PELT. The diphthong corpus is a specific corpus designed to examine the qualitative and quantitative features of British English diphthongs produced by Polish learners. It has been developed to supplement the PELT corpus on the one hand and on the other hand to answer specific phonetic questions relating to diphthongs in order to boost pronunciation training by giving precise information about diphthong-related errors and inform second language phonetics and phonology research. Both corpora consist of read speech, because at this stage in development spontaneous speech would have too many degrees of freedom. Moreover, the Polish-English Literacy Tutor is designed mainly as a reading- and pronunciation-oriented tool, so as the input to the speech-recognizer read speech was needed. In the case of a diphthong corpus, read speech was also more desirable as, by controlling for the contexts, constructed sentences used as stimuli allow for focusing on a particular pronunciation problem. Therefore, though rooted in one project, the two corpora have different levels of goals: PELT has mainly technological applications, whereas the diphthong corpus serves second language phonetics and phonology research. Both corpora could be of interest to a wider audience because of the solutions used for collecting and analyzing the second language speech data on the one hand, and a reusable format adequate for examining different phonetic phenomena on the other hand.

The platform for developing the PELT system is the Colorado Literacy Tutor, developed by the Center for Spoken Language Research (CSLR) (Cole, 1999; Cole et al., 1998, 1999, 2003) University of Colorado, Boulder, a comprehensive, scientifically-based reading program designed to teach children to read by interacting with a virtual tutor and through interactive books providing contextual feedback, reinforcement and individualized instruction. The technology developed for the Colorado Literacy Tutor involves automatic speech recognition, dialog systems and animated agents. The PELT system uses the CSLR's speech recognizer SONIC (Pellom & Hacioglu, 2003; Zhang, Pellom & Hacioglu, 2001). Recognizing second language speech, in this case Polish-accented English, is a difficult task, because of high variability of accented speech. Therefore, similar as the procedures are to the ones we applied when working on a Polish Literacy Tutor, which is system for native Polish, the PELT system aimed foreign-accented speech requires more adaptations to the specific requirements of foreign, strongly accented speech with a high degree of interpersonal variability. The Polish Literacy Tutor involved annotations of speech at the sentence level, forced-time-aligned phonetic annotation, recognizer training and defining visemes, whereas the PELT system will additionally require dealing with different characteristics of accented speech depending on the level of language proficiency of the learner, and aligning highly variable acoustic features to phonemes.

After presenting the motivational background to creating the two new Polish English corpora, comparing Polish and English phonology and describing Polglish (English with Polish accent) pronunciation, corpus collection procedures in the two cases will be presented. Next, the system of annotation, based on the previously discussed features of Polglish, will be presented. The latter will be followed by the description of the findings about Polish English pronunciation on the basis of the two corpora. Finally, the necessary next steps towards the training of the Sonic recognizer on the collected corpus of Polglish data will be discussed.

2. Motivational background

The motivational background to designing the corpora and presenting them to the public is rooted in research-oriented, technological and educational requirements of foreign language teaching and learning. PELT corpus is the first comprehensive corpus of Polish English speech, and the diphthong corpus is the first corpus of English spoken by Polish speakers designed to examine a specific phonetic phenomenon. The goal is also to integrate the corpora into a comprehensive set of high quality resources in the area of Computer Assisted Language Learning, specifically multimodal tutorial systems.

Polish-English corpora presented in this paper were needed in the context of scarcity of data based on this language combination. There are no English corpora by Polish learners in the Linguist List collection. Out of four existing corpora of English produced by Polish learners, two are devoted to written language only: PICLE (Polish sub-corpus of the International Corpus of Learner English, contains 330,000 words from over 500 essays) and PELCRA (Polish and English Language Corpora for Research and Applications) project has a Polish Learner English Corpus, which has about 500,00 words of learner data. Although another two corpora are devoted to spoken language, they are small and inadequate for examining diphthongs. The first corpus of English spoken by Poles was created by Sylwia Scheuer (1998) for the purpose of her Ph.D. thesis investigating segmental pronunciation errors by Polish learners of English and it consists of 2 hours of transcribed spontaneous speech and reading data by 13 first year English majors. The second corpus of English sentences read by Polish learners within the Polish Literacy Tutor project so far includes data from 116 learners each reading about 50 sentences. The diphthong corpus has been designed as a part of this corpus, though detailed phonetic tagging is not planned for other corpus components and the other corpus components are not controlled for segment types and contexts. Also, subjects taking part in the project recordings exhibited different levels of proficiency in English, which would make generalizations about their diphthong productions difficult. Therefore, it was decided that to realize the goal of examining diphthong productions by Polish learners, a new corpus of read-aloud, isolated sentences satisfying specific requirements was needed.

3. Examining Polish English phonetics

In Dziubalska-Kołaczyk et al. (in press) we specified the phonological features which are different in Polish and in English. They concern: rhythm, segmental inventory and devoicing. Polish is not stress-timed. As a result, vowels tend to maintain their quality and they may reduce to schwa (or be devoiced or deleted) only when phonostylistically conditioned. Other important rhythm-related differences concern: word stress – in Polish it is fixed on a penultimate syllable, and consonantal clusters – Polish is much richer in clusters in all word positions than English. Secondly, Polish is not weight-sensitive, neither in terms of vowel quality nor syllable weight. It also does not appreciate diphthongs. Thirdly, the segmental inventory of Polish is much nearer to the average balance between vowels and consonants (ca. 6 to over 20, according to Maddieson (1999)) than English. As far as system adequacy is concerned, the inventory of Polish vowels is entirely different from the English one, while in consonants, there are some important systemic as well as distributional differences. For example, 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. Finally, on a universal dimension, Polish is unmarked with reference to the process of word final obstruent devoicing, as well as interconsonantal voice agreement.

On the bases of those features we have drawn the following predictions. A Polish learner of English is predicted to have pronunciation problems stemming from all the above mentioned discrepancies between the Polish and English sound systems. The resulting 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.

An example of a typical L1-induced error is the substitution of some Polish dental or labio-dental obstruent (fricative or stop) for the English apical dental fricative. The typological rhythm difference leads, among others, to the inability to reduce unstressed vowels as well as the difficulties in stress placement. Word-final obstruent devoicing is probably the most notorious characteristics of Polglish, and predictably so, since this is a universal phonological process reinforced in Polish speakers by the system-adequacy. The above are only illustrations of the predictable mispronunciations of English by Polish learners, since a systematic survey is not possible within the scope of this paper. For the needs of PELT the most representative selection of Polglish errors has been made, with a view to sensitivize the recognizer towards those features which constitute the most perceivable traits of the foreign accent in Polish learners of English.

3.1. Predictions related specifically to diphthongs

The aim of the diphthong corpus is to provide information about the new vowel space of Polish learners of English, specifically about acoustic properties of English diphthongs produced by Poles. Polish does not have diphthongs, though it has vowel plus glide sequences comparable to English rising diphthongs, but not to centering diphthongs. Predicted are non-native properties of formant and timing relations and systematic differences in alternations applied to: simple vowels vs. diphthongs, rising vs. centering diphthongs and initial vs. final phases of diphthongs. The diphthongs containing schwa are likely to be especially difficult for Polish learners because of both qualitative and quantitative reasons. Moreover, the /ɪə/ and /eə/ diphthongs are especially likely to undergo /j/ breaking and the /ʊə/ is likely to undergo /w/ breaking. Glottal stops are likely to precede word-initial diphthongs, because they precede word-initial vowels in Polish. Vowels before nasals are heavier nasalized in Polish than in English, so it is predicted that the diphthongs before nasals are likely to be overly nasalized. English diphthong plus nasal sequences are predicted to be realized as a vowel plus semivowel plus nasal sequences. Nasal vocalization, i.e. the substitution of a nasal semivowel instead of a nasal sound occurs before fricatives in Polish, and it is predicted that this process will be transferred to English. It will be interesting to see whether the sequence in English gets realized as a vowel plus a semivowel plus a nasal semivowel plus a fricative, or whether the whole semivowel part is heavily nasalized and uniform in quality. Moreover, nasal vocalizations are predicted to appear more often in unaccented positions as these are more prone to lenitions.

4. Corpus characteristics

4.1. PELT corpus

The description of the PELT, with a different focus, corpus has already been presented in Dziubalska-Kołaczyk et al. (in press). The PELT corpus, aimed at training the recognizer to recognize accented speech depending on the level of proficiency, includes recordings of more speakers than corpora for speech recognizers dealing with native speech recognition. It has also been necessary to record learners of different levels of advancement. The number of speakers declaring a given level of advancement in English has been controlled, but since the general level of English does not closely correspond to pronunciation skills, the speakers will be divided into proficiency groups by means of statistical tests performed on the number and quality of errors they have made. The speech of any user beginning to use the program will be compared to the group characteristics and the users will thus receive pronunciation training at the appropriate level (cf. Jassem & Grygiel 2004). Corpus collection has so far been based on sentences which had been used for recording native American English speakers. These prompts were designed to ensure maximum diversity of phonetic contexts when elicited from native speakers of English, and as a result they also contained the maximum range of contexts a foreign learner might have problems with. The recording scenario so far included only read speech, as spontaneous speech at this stage of recognizer training would be too variable. Currently there are recordings of 116 speakers included in the corpus. Each speaker recorded 50 sentences, each set of sentences being different for each speaker. The speakers controlled the tempo of recordings themselves and were allowed to repeat a sentence if they wished to do so. 85 females and 31 males were recorded. Speakers' age ranged from 16 to 43, with the mean age 21,9 years and standard deviation 4,4 years. Speakers were controlled for the level of English: 24% were intermediate, 62% were advanced, and 14% were proficient learners of English. 71,6% declared to have been learning British English accent, 27,6% American English accent, and 0,9% were hesitant. Subjects were also asked to name geographical regions they came from and other foreign languages they spoke. The entire PELT corpus: sentences, labeling files and technical specifications, is approximately 3,5 GB in size and contains a total of 6032 files, corresponding to 14h 37min 37sec of running speech. The recordings were recorded, annotated and stored following EAGLES (Gibbon, Moore & Winski, 1997) and (Gibbon, Mertins & Moore, 2000), IMDI (http://www.mpi.nl/ISLE), and OLAC (http://www.language-archives.org/) recommendations. The recordings were recorded using Edirol UA 25, one channel, 24-bit resolution, and 44100 sampling frequency. Recordings were performed in a quiet office in order to obtain realistic data for tutorial system environments. A dedicated user-friendly interface was added to a simple recorder based on MCIWin functions. Audio files were checked for misreadings, repetitions etc. as they were elicited, and if necessary the speaker was asked to repeat a sentence. The PELT database was annotated by a group of students of English who completed a two year course in English phonetics. They were supposed to 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" were understood as all pronunciations accepted by educated native speakers of the standard variety of English identical to the variety declared by the subject in the interview that preceded the recording session, i.e. Received Pronunciation (RP) or General American (GA). We additionally assumed these "acceptable native readings" to be produced without disfluencies and noises. The taggers were instructed to refer to pronunciation dictionaries in the case of doubt what forms are acceptable. For the recording protocol and annotations an XML format was used. It is anticipated that the learner corpus resource will be adapted for a wide range of teaching and speaker applications.

4.2. Diphthong corpus

The corpus was designed with the requirement of including British English diphthongs (initially from the RP vowel set) in a variety of contexts. The focus of research are diphthongs as they are complex vowels, which do not exist in the Polish phoneme set. Polish is known to have vowel + semivowel sequences, but it has neither vowel duration distinctions nor vowels which are qualitatively identical to the ones appearing in English diphthongs. Therefore the research on diphthongs offers possibility of examining the interplay of substitutions of qualitative and quantitative features, and models for capturing the status of vowel plus vowel combinations (1) in English as diphthongs, and (2) in Polish as semivowel plus vowel sequences. This controlled corpus will allow examining acoustic correlates of diphthongs, properties of diphthong formants and timing, differences between rising and centering diphthongs, any specific differences in the initial and final phases of diphthongs and characteristics of timing relations in comparison to syllable structure.

The corpus design constraints were as follows. Eight British English diphthongs were taken into account: /eɪ, aɪ, ɔɪ, əʊ, aʊ, ɪə, eə, ʊə/. The conditioning criteria considered were: quality, duration, degree of nasalization, and occurrence of glottal stops before the diphthongs. There were twelve conditioning contexts in which the diphthongs were tested: word-initial, word-final, before a voiced obstruent, before a voiceless obstruent, before a nasal consonant, and before a nasal consonant followed by a fricative, and each of these conditions was tested in a stressed and unstressed position. Prosodic criteria were also taken into account and the occurrence of the words containing the examined diphthongs was controlled for stress position in a sentence, and the sentences were controlled for rhythmic units and length. All the examined diphthongs occurred in actual words embedded in sentences.

Eight diphthongs multiplied by twelve conditioning contexts give the number of 96 combinations. Due to the non-existence of real words satisfying some of the above conditions, out of potential 96 diphthong-context combinations, 29 cases could not be examined. The non-existent diphthong environments were predominantly centering diphthongs in nasal contexts. Each existent combination appeared once in the set of sentences recorded by the subjects.

30 subjects (15 females, 15 males) were recorded reading the sentences. All the subjects spoke English at an advanced level, but none of them had ever received pronunciation training or had been to an English speaking country for more than a month. All the subjects received instruction in British English.

The recording scenario involved diphthongs embedded in 61 sentences, each read three times by each subject. Subjects were instructed to read the sentences at a normal speed and in an affirmative style. The subjects controlled the tempo of recordings themselves and they were allowed to repeat a sentence when they wished to do so. The sentences were displayed on the computer screen in a random order. The recordings were collected in a quiet office environment with 22050 Hz sampling frequency and a 16-bit resolution.

The data were hand-annotated with Praat, using SAMPA phonetic alphabets for Polish and English, with an orthographic tier, and then with a tier containing segments of interest and their contexts. These were annotated with broad and narrow transcriptions and the canonical British English transcription was also noted for each word containing a diphthong of interest. Small as the corpus is, thanks to its precisely controlled parameters it is significantly informative with relation to the English diphthongs produced by Polish learners of English and its usefulness for other kinds of phonetic research or pilot studies is also foreseen. The corpus is stored in the XML format, with TASX specifications. The corpus can be obtained in a CD-ROM form by writing at abogacka@ifa.amu.edu.pl.

5. Observed pronunciation mistakes

5.1. PELT corpus

This quantitative summary reports on the analysis of 100 transcripts read and recorded by 100 subjects and tagged for errors, disfluencies and noises. The ongoing work is aimed at tagging all of the 116 transcripts corresponding to 116 speakers in the database as well as extending the speech database.

Departures from the transcript in the speech of the subjects were divided into phonetic and non-phonetic ones.

The list of phonetic errors was compiled on the basis of two empirical studies (Sobkowiak, 2005; and one by Jarosław Weckwerth, private communication). The number and type of errors to be used in the annotation of PELT was, on the one hand, a result of a compromise between the predicted discrimination and classification power of the speech recognizer trained on the data, and the pedagogical usefulness of the tool in teaching English phonetics to Polish students on the other.

There were 7277 errors found in the databse. Presented will be the percentages of ten error types referring to consonants, vowels and other errors. The ten phonetic error types have been grouped into seven major categories: five consonant error categories, one category for all vowel error types and one category for other types of errors.

The first consonant error category refered to the velar nasal /ŋ/. In Polish the velar nasal is always followed by an oral velar stop /k/ or /g/. In Polglish the following erroneous realizations of the English /ŋ/ were encountered: /ŋk, ŋg, n/ both word-finally and word-medially. The velar nasal errors amounted to 5% of all phonetic errors. The second consonant error category included voicing errors: word-final obstruent devoicing, i.e. *big* */bɪk/, and regressive assimilation of voicing in consonant clusters, e.g. *absent* */ˈæpsənt/. Errors in voicing amounted to 33,4% of all phonetic errors. The third type of errors consisted in repairing a word-final consonant cluster like /tʃt/ or /dʒd/ by inserting a schwa, e.g. *attached* */əˈtætʃət/. Such errors amounted to 0,2% of all phonetic errors. The fourth category of consonant errors referred to changing the place of articulation of a sound, e.g. /θ/ → /s/ or /t/, as in *thin* */sɪn/, and it covered 11,2% of all phonetic errors (N.B. the category did not include velar nasal errors as /ŋ/→/n/). The fifth and last consonant error category referred to changing the manner of articulation of a sound, e.g. /ʃ/→/tʃ/, as in *cliché* */ˈkliːtʃeɪ/, and it covered 0,7% of all phonetic errors.

The vowel error category covered all possible vowel error types, ranging from vowel quality and quantity errors as in *hid* */hiːd/, especially in the case of schwa as in *cater* */ˈkeɪter/, overly nasalized quality as in bend */beʊ̃nd/, monophthongization of a diphthong as in or diphthongization of a monophthong, to diphthong breaking as in *tier* */tɪʲə/. The diphthong errors amounted to 31,8% of all phonetic errors.

The category covering other phonetic errors included errors connected with stressing words and mixing varieties of English. Word stress errors related to stress placement errors as in *astronomy* */æstrəˈnɒmi/ and reduction of secondary stress, e.g. *impartially* */ɪmpəʃɪˈæli/. Inconsistent use of British and American English accent as in *after* */æftə/ instead of /ˈɑːftə/ or /ˈæftər/. 17,7% of all phonetic errors were produced in this category.

Non-phonetic departures from the acceptable native readings included word-level errors, disfluencies, restarts and noises.

Word-level errors included word deletion, word insertion, word order error, substitution of a transcript word by a different yet existent English word and misreadings – substitution of a transcript word by a different and non-existent word assuming this substitution was not motivated directly and solely by the phonetic difficulty of the transcript word but by not knowing what it means or just wrong reading of the transcript. The statistics from the total of 1478 word-level errors observed is as follows: deletions – 23,2%, insertions – 23,4%, word order errors – 0,5%, misreadings – 33,4%, substitutions – 19,5%.

Disfluencies included pauses, hesitated chunks and filled pauses. Hesitated chunks consisted of word(s) produced with hesitation, usually at a slower pace and possibly with pauses within words. The set of fillers and acknowledgements was adopted after Heeman & Allen

(1995). Out of a total of 491 disfluencies 50,1% were pauses, 39,8% were hesitated chunks and 10,1% were fillers.

There were 526 restarts in the corpus.

Noises tagged in the corpus included aside remarks, audible inhaling or exhaling, laughter, cough, throatclear, sniffing, steps, etc. There were 544 cases of noises observed.

5.2. Diphthong corpus

The corpus is still being statistically analyzed, but the qualitative analysis of the diphthongs in the corpus has been confirming predictions of the phonological system comparison and broadening the current knowledge about Polish-accented English. English diphthongs produced by Polish learners have component qualities similar to respective Polish vowel counterparts. There are no duration differences between word-final diphthongs, diphthongs before voiced obstruents and diphthongs before voiceless obstruents (N.B. syllable-final voiced obstruents are all devoiced). Sentence-initial diphthongs are preceded by glottal stops. Annotation of diphthongs in nasal contexts is especially difficult, because of widened bandwidths, which make the vocalic parts less distinctive, the lack of a vowel following a nasal, which would allow for relying on vowel formant transitions on both sides when determining the quality of the nasal, nasal releases being only optional and the first formant being often the only visible one. Diphthongs are heavily nasalized when followed by a nasal. Schwa in the nasal plus fricative context, when the nasal is vocalized, is more likely to be substituted by /o/ than in the case of non-nasal contexts. Additionally, if a nasal is followed by a fricative, in most cases the nasal is deleted, and the diphthong either remains nasalized, or just its second part is substituted by a nasal semivowel. The observed patterns of production of a diphthong plus a nasal and of a diphthong plus a nasal plus a fricative vary depending on a diphthong, but more patterns have been observed than predicted: diphthong plus a nasal, diphthong plus a nasal approximant plus a fricative, vowel plus a nasalized approximant plus a fricative, diphthong plus a nasalized approximant, vowel plus a nasalized approximant, diphthong, vowel plus a nasal. Centering diphthongs are pronounced as a vowel plus an /r/ sound, most likely because of the influence of orthography.

6. Automatic error detection

The speech corpus presented in the paper is to be 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 pronunciation error typology used to annotate the PELT corpus will constitute the basis for the preparation of accompanying acoustic models and pronunciation models. An experiment with including the errors recognized in the diphthong corpus and aligned with the reference diphthongs is also planned to test whether supplying fine-grained phonetic information can improve the performance of a second language speech recognizer. The problem of specialized models for a given type of pronunciation errors can be seen as a fine grained variant of accented speech recognition techniques described in Ikeno et al. (2003) or Kumpf & King (1996). The detector, given an acoustic observation sequence and an orthographic transcript is to evaluate the observation sequence using each of the acoustic and pronunciation models. The resulting scores for each model will 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. The implementation basis for the project is Sonic continuous speech recognition system developed at CSLR (Pellom, 2001).

7. Discussion

Because of the careful design, documentation and re-usable format, the corpora are suitable for further pilot studies research and applications in a variety of under-researched areas of Polish-English learner-performance, beyond the original goal: phonetic research including intonation research, rhythm, spectral characteristics, fluency research (variation in the number of pauses, length of pauses, ratio of pause duration to total duration time and speech rate), speech recognition evaluation, analytic work in speech synthesis and in the development of language instruction materials which are sensitive to L1-based errors. The corpora provide evidence for phonetic and phonological research as they include sounds of English in a wide variety of well-defined contexts, produced by learners at different levels of advancement. The ultimate verification of the predictions will come from implementing and using the corpora.

8. Acknowledgements

We would like to thank Wojciech Dziejma, Anna Kubisz, Paula Orzechowska and Paulina Zydorowicz who collected most of the data and greatly helped in the annotation. Thanks go to Włodzimierz Sobkowiak for providing his PDI (Phonetic Difficulty Index) and the list of Polglish errors. Jarosław Weckwerth shared with us his observations on the degree of difficulty of particular features. None of this would have happened without Ron Cole's encouragement and assistance, and generous sharing of the CSLR technology with our team.

Research on PELT was supported by Adam Mickiewicz University grant.

9. References

Cole R. (1999). Tools for research and education in speech science. In *Proceedings of the International Conference of Phonetic Sciences*. San Francisco, CA, pp. 1277-1280.

Cole R., T. Carmell, P. Connors, M. Macon, J. Wouters, J. de Villiers, A. Tarachow, D. Massaro, M. Cohen, J. Beskow, J. Yang, U. Meier, A. Waibel, P. Stone, G. Fortier, A. Davis, and C. Soland. (1998). Intelligent animated agents for interactive language training. In *STiLL: ESCA Workshop on Speech Technology in Language Learning*. Stockholm, Sweden.

Cole R., Massaro, D. W., de Villiers, J., Rundle, B., Shobaki, K., Wouters, J., Cohen, M., Beskow, J., Stone, P., Connors, P. Tarachow, A., & Solcher. D. (1999). New tools for interactive speech and language training: Using animated conversational agents in the classrooms of profoundly deaf children. In *Proceedings of ESCA/SOCRATES Workshop on Method and Tool Innovations for Speech Science Education*. London, UK, pp. 45-52.

Cole, R.A., Van Vuuren, S., Pellom, B., Hacioglu, K., Ma, J., Movellan, J., Schwartz, S.,Wade-Stein, D., Ward, W. and Yan, J. (2003). Perceptive Animated Interfaces: First Steps Toward a New Paradigm for Human–Computer Interaction. In *Proceedings of the IEEE: Special Issue on Human-Computer Multimodal Interface*, 91 (9), pp. 1391-1405.

Dziubalska-Kołaczyk, K., Bogacka, A., Pietrala, D., Wypych, M., Krynicki, G. (In press). PELT: An English language tutorial system for Polish speakers. In *Proceedings of the MultiLing Conference*. Stellenbosch, South Africa.

Gibbon, D., R. Moore and R. Winski (eds.). (1997). *Handbook of Standards and Resources for Spoken Language Systems*. Berlin: Mouton de Gruyter.

Gibbon, D., I. Mertins and R. Moore. (2000). *Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Product Evaluation*. Dordrecht: Kluwer Academic Publishers.

Heeman, Peter A. and James Allen. (1995). The Trains 93 Dialogues. *TRAINS Technical Note*, pp. 94-2.

Ikeno, A., Pellom, B., Cer, D., Thornton, A., Brenier, J., Jurafsky, D., Ward, W., Byrne, W. (2003). Issues in Recognition of Spanish-Accented Spontaneous English. In *ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, Tokyo, Japan.

Jassem, W. , Grygiel, W. (2004). Off-line classification of Polish vowel spectra using artificial neural networks. *Journal of the International Phonetic Association*, 34 (01), pp. 37-52.

Kumpf, K., King R.W. (1996). „Automatic Accent Classification of Foreign Accented Australian English Speech”. In *Proceedings of ICSLP 1996*. Philadelphia, USA. pp 1740—1743.

Maddieson, I. (1999). In search of universals. *ICPhS99.* vol. 3. 2521-2528.

Pellom, B. (2001). “SONIC: The University of Colorado Continuous Speech Recognizer”. University of Colorado, Technical Report #TR-CSLR-2001-01, Boulder, Colorado.

Pellom, B., Hacioglu, K.. (2003). Recent Improvements in the CU Sonic ASR System for Noisy Speech: The SPINE Task. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*. Hong Kong.

Sobkowiak, W. (2005). Phonolapsological equivalence and similarity in the English lexicon. In F.Kiefer, G.Kiss & J.Pajzs (eds). *Papers in computational lexicography, Proceedings of the 8th International Conference on Computational Lexicography (COMPLEX 2005),* Budapest: Linguistics Institute. 200-212.

Zhang, J., Ward, W., Pellom, B., Yu, X., Hacioglu, K. (2001). Improvements in Audio Processing and Language Modeling in the CU Communicator. In *Proceedings of Eurospeech 2001*. Aalborg Denmark.