Development and Evaluation of Polish Speech Corpus for Unit Selection Speech Synthesis Systems

[1]G. Demenko, [1]J. Bachan, [2]B. Möbius, [1]K. Klessa, [3]M. Szymański, [3]S. Grocholewski

[1]Institute of Linguistics, Department of Phonetics, Adam Mickiewicz University, Poznań
[2] Institute of Natural Language Processing, University of Stuttgart, and
Institute of Communication Sciences, University of Bonn
[3]Institute of Computing Science, Poznań University of Technology.
lin@amu.edu.pl, jolabachan@gmail.com, bernd.moebius@ims.uni-stuttgart.de, klessa@amu.edu.pl, marcin.szymanski@cs.put.poznan.pl, stefan.grocholewski@cs.put.poznan.pl

Abstract

This paper presents the results of a set of experiments assessing the perceived quality of the Polish version of the BOSS unit selection synthesis system. The experiments aimed to evaluate the potential improvement of synthesis quality by three factors pertaining to corpus structure and coverage as well as levels of corpus annotation. The three factors affecting synthesis quality were (i) manual vs. automatic corpus annotation, (ii) coverage of CVC triphones in rich intonational patterns, and (iii) coverage of complex consonant clusters. Results indicate that a manual correction of automatic annotations enhances synthesis quality. Increased coverage of CVC sequences and consonant clusters also improved the perceived synthesis quality, but the effect was smaller than anticipated.

Index Terms: speech synthesis, unit selection, corpus validation, perceptual evaluation

1. Introduction

Speech synthesis systems are based on machine learning techniques and rely heavily on training a speech material representative of a specific task. The quality of the synthesised speech depends on the text type and synthesis domain: intonation is very natural for restricted domain, e.g. news or weather forecast, and prosodically stable speech (read or dictated texts) which is distinguished by quite flat intonation, stable voice quality and easily predictable duration of the speech units. Ideally, the speech segments should cover all phonetic variations, all prosodic variations, and all speaking modes. Due to the limited speech material to be recorded per speaker the focus has to be on the coverage of phonetic and prosodic variations which means that these speaking modes should be quite uniform over the domains chosen.

In order to meet the requirements concerning the coverage of segmental and suprasegmental features, the size of databases for speech technology purposes is expected to be substantial, e.g. according to ECESS guidelines [1, 2] the overall duration of the recorded speech signals for speech synthesis database should be approximately ten hours. Applying corpora of this size is reported to give comparably good results also for Slavic languages, e.g. for Czech [cf. 3]. The large corpus size may however entail problems with further processing even considering the progressing development of the tools enabling automatic transcription and segmentation and the results provided by them are more and more convincing, however it is still often the case that manual inspection proves indispensable.

For the purpose of this study it was decided to test a smaller, low-redundant database using its various subsets to observe the database structure influence on the quality of the resulting synthesised speech.

2. Polish Speech Corpus

2.1. Corpus contents and structure

The unit selection corpus of the Polish BOSS synthesizer (see section 3) comprises 115 min. of speech read by a professional radio speaker during several recording sessions, supervised by an expert phonetician. Since the problem of constructing an effective low redundant database for flexible concatenative speech synthesis has not been solved satisfactorily either for Polish or any other language, we have decided to use various speech units from different mixed databases as follows:

- Base A: Phrases with most frequent consonant structures. Polish language has a number of difficult consonant clusters. 258 consonant clusters of various types were used.
- Base B: All Polish diphones realised in 92 grammatically correct but semantically nonsense phrases.
- Base C: Phrases with CVC triphones (in non-sonorant voiced context and with various intonation patterns). 664 phrases were recorded for triphone coverage.
- Base D: Phrases with CVC triphones (in sonorant context and with various intonation patterns). The length of the 985 phrases varied from 6 to 14 syllables to provide coverage of suprasegmental structures.
- Base E: Utterances with 6000 most frequent Polish vocabulary items. 1109 sentences were recorded.
- Base TEXT: 15 minutes of prose and newspaper articles.

2.2. Segmental and suprasegmental annotation of the corpus

The computer coding conventions were drawn up in SAMPA for Polish created by J. C. Wells [4] with revisions and extensions, and in the IPA alphabet [5, 6]. Two sets of characters were precisely defined for the exact grapheme-to-phoneme mapping for the Polish language – an input set of orthographic characters and an output phonetic/phonemic alphabet. An inventory of 39 phonemes was employed for broad transcription and a set of 87 allophones was established for the narrow transcription of Polish. The speech data were first transcribed with PolPhone [7] and segmented automatically with SALIAN software [8]. Then, the label files were manually corrected by a team of phoneticians on the basis of perceptual and acoustic analyses of the speech signal.

The recording files are annotated on both segmental and suprasegmental level thus respecting the requirements of the annotation system for unit selection: the information on phone, syllable and word boundaries, syllable stress, phrase boundaries of different type and strength. For prosody modeling, only the fundamental types of prosodic structures were distinguished, such as word stress and phrase accent placement, accent type or phrase boundary type according to the BOSS Label File (BLF) format [cf. 9]. On the phrase level information about sentence and intonation type was provided. On the syllable level pitch accent types were marked. Pitch accents are determined by pitch variations occurring on the successive vowels/syllables and pitch relations between syllables. The annotation of pitch accent types can be complex because it may include combinations of many acoustic features (e.g. pitch movement direction, range of the change of pitch, pitch peak position). With a view to simplifying the annotation of the pitch accents we took into consideration only two features: direction of the pitch movement and its position with respect to accented syllable boundaries. The resulting inventory of pitch accent labels was described e.g. in detail in [10, 11]. For establishing syllable boundaries for Polish, the rules based on 20-million word Polish lexicon were set by an expert (based on the assumed relationship between *sonority* and *syllable structure*) [12]. The syllabification rules were fully automatised and implemented in software program created specifically for the purpose: *Annotation Editor*, which integrates also Wavesurfer [13] and SALIAN [8], stress and accent control unit, and a Windows editor of BOSS Label Files.

3. Implementation of Polish TTS modules in Bonn Open Speech Synthesis System

Two Polish modules were implemented for BOSS so far [cf. 14]: the duration prediction module and the cost functions module. In BOSS, cost functions may be effective on both nodes and arcs (representing speech units and concatenations, respectively) of the network of candidate units. Currently, the node cost function applied in the Polish version of BOSS consists of the following components:

- the absolute difference between the CART-predicted segment duration and the candidate unit duration (in ms),
- the boolean difference between predicted and actual stress value, multiplied by 10,
- the discrepancy regarding phrase type (question or statement, raising or falling intonation) and phrase location within a sentence (final or comma-terminated), multiplied by 20.

In the most recent implementation, two features are considered by the transition cost function:

- the Euclidean MFCC distance between the left segment right edge and the right segment left edge,
- the absolute F0 difference, analogously (currently only for phone segments).

The auditory experiments suggest that relocation of the syllable within the phrase should be particularly penalised.

Several experiments to predict segmental duration with CART were carried out, using various sub-corpora of the speech database. The best obtained results (the overall correlation of 0.8) were reported in Klessa [15] .

4. Evaluation Results

After preliminary evaluation of the speech output [11] and diagnostic annotation evaluation with the use of an Automatic Close Copy Speech (ACCS) synthesis tool [16] as an audio screening procedure, the BOSS synthesis system for Polish was assessed in five speech quality judgement tests based on SAM/EAGLES standards [17]. The assessment was administered to 20 Polish students of different philologies, 11 females and 9 males. The subjects were from 20 to 25 years old. They took the tests separately and they listened to the stimuli via headphones to minimalise the ambient noise. The tests were self-paced and the stimuli could be played as many times as the subjects required. Each session lasted 50-90min.

Three factors affecting synthesis quality were investigated by means of a total of 5 perception tests:

- Factor 1: Improvement of quality of synthetic speech based on manually corrected annotations. Test 1a: preference test; Test 1b: mean opinion score (MOS).
- Factor 2: Improvement of quality of synthetic speech where the unit selection system contains a database rich of phrases with CVC triphones in sonorant context and various intonation patterns. Test 2a: preference test; Test 2b: mean opinion score (MOS).
- Factor 3: Improvement of quality of synthetic words with complex consonant clusters where the unit selection system contains a database rich of complex consonant clusters. Test 3: preference test.

In the preference tests the material was divided into four groups, each comprising 25% of the total number of stimuli: AB pairs, BA pairs, AA pairs, BB pairs. "A" represents the stimuli synthesised using the full manually prepared inventory of speech. These stimuli are called *standard* in the paper. "B" stands for the stimuli synthesised for the examination of the three factors of the BOSS speech corpus described above.

The subjects of the tests could choose among three answers: "A" sounds better; "B" sounds better; both sounds are the same.

The identical pairs AA and BB were added as noise to count the false alarm rate which indicates how many pairs of identical stimuli are incorrectly recognised as including different stimuli [18]. If the false alarm rate was above 30% then the subjects' results were excluded from the statistical analyses in this research.

4.1. Factor 1: Manual vs. automatic annotations

Factor 1 aimed to examine the influence of the manual corrections of the automatically generated database annotations. The first set of test stimuli consisted of the synthetic sounds synthesised in the standard way ("A" stimuli). The second set of stimuli was generated based on the automatic annotations ("B" stimuli).

The database annotation files consist of: segmental phonetic transcription; syllabic stress information; and phone, syllable and phrase level segmentation. For the perception tests, all of the above were annotated automatically, whereas in the standard synthesis setup all annotations were checked and, if necessary, corrected by an expert.

For the tests, sentences from three different groups were used: (i) sentences from the BOSS corpus, e.g. '*Bezmięsny poniedziałek jest katastrofą.*' ('Vegetarian Monday is a catastrophe.'); (ii) one-phrase sentences containing common vocabulary, e.g. '*Nie mam teraz ochoty na lody*.' ('I don't feel like having ice creams right now.'); and (iii) multiple-phrase sentences containing difficult and rare vocabulary, e.g. '*Sposób w jaki państwo angażuje się w Afganistanie, do złudzenia przypomina początki operacji irackiej sprzed trzech lat*.'('The way in which the country gets involved in Afghanistan is the spitting image of the beginnings on the Iraq operation from three years ago.')

.

4.1.1. Test 1a: preference test

60 utterances were used in the preference test, viz. 30 sentences from the *standard* synthesis corpus ("A" stimuli) and 30 sentences based on the automatically generated annotations ("B" stimuli). The results of Test 1a are presented in Table 1.

Table 1: *Results of the preference tests (in percent): Test 1a, Test 2a and Test 3. N = number of subjects; "none" indicates that the subject did not choose any of the three possible answers.*

Test	Group	N	A	B	same	none
Test 1a	Male	8	38.33	26.67	35.00	0.00
	Female	9	36.67	30.37	32.59	0.37
	Overall	**17**	**37.50**	**28.52**	**33.80**	**0.19**
Test 2a	Male	8	30.47	22.27	47.27	0.35
	Female	9	27.08	25.35	47.22	0.00
	Overall	**17**	**28.78**	**23.81**	**47.24**	**0.17**
Test 3	Male	8	23.75	13.13	63.13	0.00
	Female	11	24.09	9.77	65.91	0.23
	Overall	**19**	**23.92**	**11.45**	**64.52**	**0.11**

The results show that subjects chose the *standard* stimuli as sounding better most often. This suggests that the manual corrections of the automatic annotations improve the quality of the synthetic speech in the BOSS system for Polish.

4.1.2. Test 1b: MOS test

For the MOS test 30 utterances were used: 15 *standard* ("A") stimuli and 15 "B" stimuli. The tests results are presented in Table 2.

The results show that the sentences which were present in the BOSS corpus were assessed better when they were synthesised in the *standard* way. However, utterances containing common vocabulary generated on the automatic annotations received better scores, which shows that the used tool for automatic annotation are very good. When it comes to the utterances with difficult vocabulary, the scores were low and similar for both types of stimuli. Taken together, synthetic stimuli based on manually corrected annotations received higher scores.

Table 2: *Test results for Test 1b. N = number of subjects, MOS score (scale 1-5), MOS score standard deviation (S.D.).*

Group	N	MOS		MOS S.D.	
		A	B	A	B
From Corpus (10)					
Male	9	3.36	2.67	2.25	1.70
Female	11	2.89	2.31	2.18	1.71
Overall	**20**	**3.10**	**2.47**	**2.19**	**1.71**
Common Vocabulary (10)					
Male	9	3.16	3.29	2.05	2.19
Female	11	2.71	3.02	1.87	2.14
Overall	**20**	**2.91**	**3.14**	**1.92**	**2.15**
Difficult Vocabulary (10)					
Male	9	1.96	1.98	1.32	1.38
Female	11	1.51	1.47	1.14	1.06
Overall	**20**	**1.71**	**1.70**	**1.20**	**1.17**
Overall (30)					
Male	9	2.82	2.64	1.99	1.86
Female	11	2.38	2.27	1.87	1.79
Overall	**20**	**2.58**	**2.44**	**1.89**	**1.80**

4.2. Factor 2: Phonetically rich phrases

Factor 2 aimed to investigate the improvement of the quality of synthetic speech in a context in which the unit selection system comprises a database rich of phrases with CVC triphones in sonorant context and with various intonational patterns.

For the purpose of comparison, the part of the database ("Base D") that includes this kind of phonetically rich phrases was switched off during synthesis of type "B" stimuli. Therefore, type "B" stimuli were generated using a smaller corpus: 985 utterances were excluded, resulting in a corpus of 88 min of speech.

For testing factor 2, sentences from four different groups were used: (i) sentences from Base D, e.g. '*Czy to jest wyraz* ***promil****, czy* ***promień****?*'('Is it a word per mill or a radius?'); (ii) sentences containing keywords from Base D, e.g. '*Zaśpiewasz to w tonacji* ***bemol****?*' ('Will you sing it in flat major?'); (iii) sentences containing common vocabulary, e.g. '*Jem obiad w kuchni przy stole.*' ('I eat dinner in the kitchen at the table.'); and (iv) semantically unpredictable sentences, e.g. '*Jezioro tańczyło o myszy wysoko nad ziemią.*'('A lake danced about a mouse high above the ground.')

4.2.1. Test 2a: preference test

In the preference test 64 utterances were used: 32 *standard* stimuli and 32 "B" stimuli synthesised when Base D was excluded. The results of Test 2a are presented in Table 1.

The results show that the *standard* stimuli ("A") were generally preferred over B-type stimuli. However, almost 50% of the pairs AB and BA were assessed as sounding the same. This suggests that there is little difference between the stimuli A and B and therefore Base D does not improve the quality of speech very much.

4.2.2. Test 2b: MOS test

In the MOS test 32 utterances were used: 16 *standard* stimuli and 16 synthesised without Base D. The results of Test 2a are presented in Table 3.

The results show that the quality of the synthetic speech is assessed poorer when Base D is not used for the synthesis. However, it has to be pointed out that Base D comprises 28 min of speech and the synthesis with the standard setup of the BOSS system for Polish contains 115 min of recorded speech. This means that for this analysis 23% of the BOSS corpus was excluded and the B-type stimuli were generated on a corpus of only 88 min of speech.

The quality of the semantically unpredictable sentences was assessed as being the worst of the scenarios tested, and the overall results for stimuli types "A" and "B" were approximately equal. The conclusion may be drawn that a database of recordings containing CVC triphones does not boost the quality of the speech output very much.

Table 3: *Results for Test 2b (columns as in Table 2).*

Group	N	MOS		MOS S.D.	
		A	B	A	B
From Base D (10)					
Male	9	3.44	2.64	2.14	1.73
Female	11	3.38	2.25	2.18	1.51
Overall	**20**	**3.41**	**2.43**	**2.16**	**1.59**
Keywords from Base D (10)					
Male	9	3.49	2.89	2.15	1.90
Female	11	3.00	2.58	1.97	1.80
Overall	**20**	**3.22**	**2.72**	**2.04**	**1.85**
Common Vocabulary (6)					
Male	9	3.11	3.07	2.15	1.92
Female	11	2.61	2.42	1.90	1.78
Overall	**20**	**2.83**	**2.72**	**2.01**	**1.82**
Meaningless sentences (6)					
Male	9	2.00	1.93	1.47	1.37
Female	11	1.67	1.73	1.18	1.13
Overall	**20**	**1.82**	**1.82**	**1.30**	**1.21**
Overall (32)					
Male	9	3.13	2.67	2.12	1.82
Female	11	2.79	2.29	1.99	1.64
Overall	**20**	**2.94**	**2.46**	**2.03**	**1.70**

4.3. Factor 3: Complex consonant clusters

The Polish language is famous for allowing complex consonant clusters, which occur in many common words. As many as 6 consonants may occur in one cluster, such as /mpstʃj/ in '*o przestępstwie*'('about a crime'). For the Polish BOSS system a database of recordings containing words with the most frequent complex consonant clusters was recorded. This database is called "Base A" in the corpus.

The test was designed to examine the quality of the synthesis of complex consonant clusters when they are concatenated from bigger units vs. complex consonant clusters based on concatenation of single phones. For the purpose of studying this factor, Base A was excluded from the synthesis to create B-type stimuli. Altogether, 285 utterances (12 min of speech) were excluded from the corpus.

In this test, single words containing complex consonant clusters were used. The words were taken from Base A and included 4-5 consonants in a cluster. Some of the words used in the test with their clusters and the information about the respective results of unit selection are presented in Table 4. The results show that for the synthesis of A-type stimuli many units in consonantal context from Base A were used. Since for the synthesis of the A-type stimuli the system chose the units from the Base A as the best candidates, it may be anticipated that for the synthesis of the B-type stimuli the quality of the synthesised words will be lower.

4.3.1. Test 3: preference test

80 stimuli were included in the test: 40 *standard* words and 40 words synthesised without the recordings from Base A ("B" stimuli). The results of Test 3 are presented in Table 1.

The results show that the *standard* stimuli (type "A") were assessed better overall. This indicates that the database containing complex consonant clusters indeed improves the quality of speech output. However, considering the large proportion of stimuli pairs A and B judged with no preference of one over the other (65%), the advantage of including complex clusters in the recorded database was not directly confirmed.

Table 4: *Example words from test 3. Column 1 – orthographic notation; Column 2 – SAMPA transcription with the target clusters in bold; Column 3 – segments selected for synthesis from the file in which the keyword occurred, with their right and left context; Column 4 – 2-or-more-phone segments taken from other files, the letters in brackets stand for the Base containing the file.*

Column 1	Column 2	Column 3	Column 4
z marchwią	zma**rxfj**ow~	r-xf-j	z-ma-r (A)
rozstrojone	ro**sstr**ojone	s-stro-j	r-os-s (A); o-jo-n (D)
westchnienie	ve**stxn'**en'e	t-x-n'	
przedwstępny	pSe**tfst**empny	S-etf-s	s-te-m (A)
odświeżyć	o**ts'fj**eZyt^s'	t-s'-f	s'-fj-e (A)
wzbierać	**vzbj**erat^s'		e-ra-t^s' (A)

5. Discussion and Conclusion

The results of the quality tests showed that it is sufficient to use a 115 min database to obtain a relatively good synthetic speech of a neutral style. Such database may be based mostly on read speech recordings where the intonation is generally rather flat and significant intonation changes occur in the phrase final and phrase initial positions.

The evaluation tests confirmed the expected improvement of the synthesis where the automatic annotation was manually corrected by trained phoneticians. However, the results also showed that the applied tools for the automatic annotation performed well. Currently, software is being developed which will enhance the automatic segmentation engine and include automatic annotation of intonation structures.

When it comes to the usefulness of including phonetically rich phrases in the unit selection synthesiser, it was confirmed that such a database boosts the quality of the synthetic speech. It should be noted however that it is quite possible that a reduction of the database size by 23% contributes to the reduced perceived quality of the synthesis. The effects of the database size and its composition are thus likely to be confounded, and their relative contribution to perceptual quality reduction is difficult to quantify at this stage.

Similar shift of quality might be expected when a part of the database was excluded for the experiment with the consonant clusters. Analysis of the selection procedure showed that, indeed, the system promotes the candidates originally occurring within consonant clusters.

As regards the technical solutions for the synthesis system it is planned to further develop the cost function and implement a more sophisticated prosody control module. Another necessary improvement is needed for the concatenation method and join costs.

With respect to the annotation techniques, it is intended to create tools enabling full automatisation of both segmental and suprasegmental annotation of Polish speech data for the needs of unit selection. The work is going on on developing tools for annotation of expressive speech.

The database will be elaborated in two respects: first, for neutral speech synthesis improvement - new recordings of read speech will be provided using the text material covering approximately 10,000 Polish triphones in syntactically and phonetically rich sentences (prepared within the present project); second, the database will be expanded for expressive speech.

Finally, it is planned to verify specifications of our speech corpus structure with ECESS guidelines and submit the database for validation and expertise by an external institution, for example ELDA [19].

6. Acknowledgements

This research was supported by the Polish Ministry of Scientific Research and Information Technology, project no. R00 035 02. It has also been supported by an Alexander von Humboldt Polish Honorary Research Fellowship awarded to one of the authors, Bernd Möbius.

7. References

[1] Bonafonte, A., Höge, H., Tropf, H. S., Moreno, A., Heuvel, H., Sündermann, D., Ziegenhain, U., Pérez, J. and Kiss, I. "TC-STAR Deliverable no.: D8. TTS Baselines and specifications." 2005. Online: http://suendermann.com/su/pdf/tcstar2005b.pdf, accessed on 13 Apr 2008.

[2] ECESS: European Center of Excellence on Speech Synthesis Web Page. Online: http://www.ecess.eu/, accessed on 13 Apr 2008.

[3] Romportl, J. and Kala, J., "Prosody Modelling in Czech Text-To-Speech Synthesis", 6th ISCA Workshop on Speech Synthesis (SSW-6) Proc., Bonn, 2007.

[4] Wells, J. C., "SAMPA Computer Readable Phonetic Alphabet". Online: http://www.phon.ucl.ac.uk/home/sampa/polish.htm, accessed 10 Apr 2008.

[5] The International Phonetic Association (IPA) Homepage. http://www.arts.gla.ac.uk/ipa/ipa.html, accessed on 13 Apr 2008.

[6] Jassem, W. "Illustrations of the IPA. Polish." Journal of the International Phonetic Association, vol. 33 (1): 103-107, 2003.

[7] Demenko, G., Wypych, M., and Baranowska, E., "Implementation of Grapheme-to-Phoneme Rules and Extended SAMPA Alphabet in Polish Text-to-Speech Synthesis", Speech and Language Technology, vol. 7. [Ed.] PTFon, 79-97, Poznań, 2003.

[8] Szymański, M. and Grocholewski, S., "Semi-Automatic Segmentation of Speech: Manual Segmentation Strategy. Problem Space Analysis", Advances in Soft Computing, Computer Recognition Systems, 747-755, Springer Berlin, 2005.

[9] Breuer, S. and Abresch, J. "Phoxsy: Multi-phone Segments for Unit Selection Speech Synthesis", International Conference on Spoken Language Processing (ICSLP) Proc., Jeju, 2004.

[10] Demenko, G. and Wagner, A., "Analysis of accented syllables in different prosodic contexts for use in Unit Selection Speech Synthesis", Archives of Acoustics, vol.30. 2006.

[11] Demenko, G., Klessa, K., Szymański, M., Bachan J., The design of Polish Speech Corpora for expressive speech synthesis in BOSS system, PPEE Conference. 2007.

[12] Śledzinski, D. "Charakterystyka zbitek spółgłoskowych języka polskiego w aspekcie profile sonorności oraz liczebności," Speech and Language Technology, vol. 8. [Ed.] PTFon, 65-74, Poznań, 2006.

[13] Sjölander, K. and Beskow, J., Wavesurfer Homepage. Online: http://www.speech.kth.se/wavesurfer/, accessed on 13 Apr 2008.

[14] The Bonn Open Synthesis System (BOSS). Online: http://www.ifk.uni-bonn.de/forschung/abteilung-sprache-und-kommunikation/phonetik/sprachsynthese/boss, accessed on 14 Apr 2008.

[15] Klessa, K., Szymański, M., Breuer, S. and Demenko, G., "Optimization of Polish Segmental Duration. Prediction with CART", 6th ISCA Workshop on Speech Synthesis (SSW-6) Proc., Bonn, 2007.

[16] Gibbon, D. and Bachan, J., "An automatic close copy speech synthesis tool for large-scale speech corpus evaluation" in LREC 2008 Proc., Marrakech, 2008 (forthcoming).

[17] Gibbon, D., Moore, R. and Winski, R. Handbook of Standards and Resources for Spoken Language Systems. Berlin: Mouton de Gruyter, 1997.

[18] Schneider, K., Lintfert, B., Dogil, G. and Möbius, B. "Phonetic grounding of prosodic categories", in S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, and Johannes Schließer [Eds], Methods in Empirical Prosody Research, 335-361. Walter de Gruyter, Berlin, 2006.

[19] ELDA: Evaluations and Language resources Distribution Agency. Online: http://www.elda.org/, accessed 13 Apr 2008.