Prosodic Analysis of German Produced by Russian and Chinese Learners

Andreas Hilbert [1], *Hansjörg Mixdorff* [1], *Hongwei Ding*[2], *Hartmut R. Pfitzinger*[3] *and Oliver Jokisch*[4]

[1] Department of Informatics and Media, Beuth University of Applied Sciences, Berlin, Germany
[2]School of Foreign Languages, Tongji University, Shanghai, China
[3]Inst. of Phonetics and Digital Speech Processing, Christian-Albrechts-University Kiel, Germany
[4] Institute for Acoustics and Speech Communication, TU Dresden, Germany

hilbert|mixdorff@beuth-hochschule.de, hongwei.ding@tongji.edu.cn, hpt@ipds.uni-kiel.de,
oliver.jokisch@tu-dresden.de

Abstract

This study compares utterances by Russian and Chinese learners of German with those of native speakers. Adopting the methodology of an earlier study on accented English, we rated the utterances for strength of foreign accent. We aim to find measurable prosodic differences accounting for the perceptual results. Our outcomes indicate, inter alia, that unaccented syllables are relatively longer compared with accented ones in the Chinese data than in the Russian and German data. Furthermore, the inter-speaker-correlations of syllabic durations in utterances of one and the same sentence are much higher for German speakers than for Russian and Chinese learners of German. Russian speakers tend to use a larger range of *F0* and produce more pitch-accents than German speakers.

Index Terms: foreign accent, prosodic analysis

1. Introduction

Although foreign accent is mostly associated with segmental deviations from a native norm, prosodic differences certainly account for many difficulties in understanding accented speech (see, for instance, [1][2][3]). In the current study we examine speech collected from Russian and Chinese learners of German. In the line of earlier work on Australian English the data was assessed by native listeners for strength of foreign accent on a scale from 1 to 5 [4][5]. We attempt to perform a prosodic analysis of the recordings and compare them with corresponding utterances by native German speakers in order to establish objective parameters that best reflect foreign accent, as well as are correlated with the subjective measures of foreign accent. Whereas German is often classified as a stress-timed language Russian and Chinese are regarded a syllable-timed, the latter being a tone language, contrasts which obviously pose a number of prosodic problems for learners of German.

2. Speech Material and Method of Analysis

The complete corpus consists of recordings from readings of 11 target sentences. Eight of these are part of the PHONDAT corpus and three from the VEITH corpus [6]. The sentences were uttered by a total 15 Russian (7 male, 8 female) and 17 Chinese (9 male, 8 female) learners of German. In addition, the corpus contains recordings by 14 native German (9 male, 5 female). Most of the data were recorded in a sound-treated room at a sampling rate of 16 kHz at 16 bits.

In a first step, all recordings were forced-aligned on the word and phone-levels using the *LINGWAVES* German Forced Aligner [7]. The text targets were the expected results of the reading task. Examination of alignment results, however, showed that the procedure frequently produced errors and mismatches, as well as pause insertions. Other problems concerned hesitations and repairs.

Ultimately a total of 364 utterances were selected for further analysis, 161 from Chinese speakers, 86 from Russian and 117 from German speakers (4515 syllables, 12157 phones) that contained the desired target sentences. These will henceforth be referred to as *CN*, *RU* and *DE* sets, respectively.

Since we were primarily interested in the intonational and rhythmic properties of the accented speech, the forced alignment was re-run on the syllabic level by editing the phonetic transcriptions yielded from the text-based alignment in the first step to employ syllabic subdivisions.

Subsequently, the label files from the alignment procedure were converted to *PRAAT* TextGrid format [8] and combined in a single TextGrid containing word, syllable and phone labels. The syllabic boundaries were then hand-corrected and phone labels automatically adjusted in proportion to the syllable. At this stage we were not interested in the identity and exact boundaries of phones actually realized, but the rhythmic structure of the utterances. In the *CN* data, however, we found many instances of epenthetic vowels which in some cases alter the rhythmic structure and therefore need to be taken into account.

Since foreign accent often involves wrong accent placement, accented syllables were identified perceptually, and marked as appropriate or inappropriate with respect to the underlying word, as well as with regard to the German default sentence intonation. Another known issue is inappropriate marking of phrase boundaries, that is, for instance, the occurrence of falling boundary tones where rising ones are required. Hence all phrase endings were examined as to the type of boundary tone associated with them.

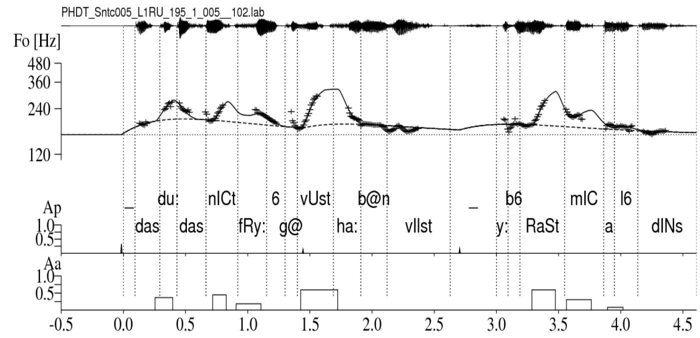

Figure 1: *Example of analysis of sentence "Dass du das nicht früher gewusst haben willst, überrascht mich allerdings."-" (The fact) that you claim not to have known this earlier surprises me." uttered by a female Russian speaker.*

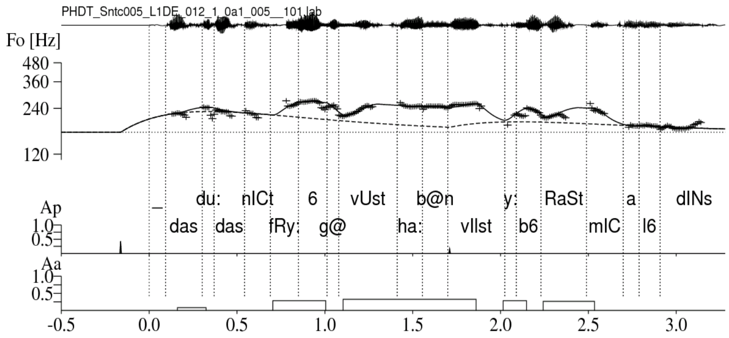

Figure 2: *Example of analysis of the same sentence as in Figure 1, uttered by a female German speaker. The phrase boundary after "willst"-[vIlst] is associated with a high boundary tone, in contrast to the same sentence uttered by a Russian speaker (Figure 1).*

In order to compare the intonational properties of the three data sets, *F0* values were extracted at a step of 10*ms* using the *PRAAT* default pitch extraction settings.

All utterances were subjected to Fujisaki model [9] parameter extraction as shown in Figure 1 (produced by a female Russian speaker).

The figure displays from the top to the bottom: The speech wave form, the *F0* contour (+signs: extracted, solid line: model-based), the SAMPA transcription of the underlying phrase and accent commands.

The perceptual local speech rate (PLSR) is a psychophysical measure which was developed [10] because earlier measures such as the local syllable rate and the local phone rate are not well-correlated, meaning that they represent different aspects of speech rate. Perception experiments with short stretches of speech being judged on a rate scale revealed that neither syllable rate nor phone rate is sufficient to predict the perception results. Subsequently it was shown that a linear combination of the two measures yielded a correlation of r=0.91 and a mean deviation of 10% which is accurate enough to successfully extract PLSR from large spoken language corpora. The result is a smooth contour of local speech rate values aligned with the speech signal where a value of 100% represents a typical average speech rate while 50% being approx. half of it and 200% being roughly twice the average.

3. Perceptual Rating

16 native speakers of German (10 male, 6 female) most of whom were second year students at Beuth University took part in a perceptual rating test. To that effect they were given questionnaires in the form of a *Winword* document containing hyperlinks to the sound files. We included all *CN* and *RU* data and 28 utterances from the *DE* set as a reference. Participants were given a list of the utterances to be rated containing the text of each sentence. They were asked to (1) rate their overall impression of the strength of foreign accent on a scale from 1-5, with 1 being native-like, 5 a very heavy accent, (2) rate the prosodic quality, also on a scale from 1-5 and (3) mark or rewrite words in the text they found hard to understand. They were requested to listen to the stimuli for a maximum of three times. We found that inter-rater correlation of judgments was above 0.7 in all cases. Split correlation for two groups is .953.

Table 1 displays ratings for the three different stimuli sets.

All learners had been classified by their teachers as belonging to either beginner, intermediate and advanced levels. We therefore calculated the mean perceptual ratings for each of the levels. Table 2 shows that productive performance improves with the time learners spend, but also that the Chinese learners on comparative levels perform much more poorly than their Russian counterparts. We have to bear in mind that individual results may vary considerably. Still it is surprising that the mean rating difference between the beginner and the advanced levels is only .88 for Russians and .52 for Chinese.

Table 1: *Mean and standard deviation of perceptual ratings for German native, Russian and Chinese speakers.*

set	global rating	prosodic rating
DE	1.09/.20	1.39/.19
RU	2.47/.57	2.38/.57
CN	3.28/.52	3.33/.53

Table 2: *Mean and standard deviation of perceptual ratings for beginner, intermediate and advanced levels.*

level	set	global rating	prosodic rating
beginner	*RU*	2.89/.36	2.80/.52
	CN	3.54/.46	3.65/.50
intermediate	*RU*	2.51/.43	2.31/.42
	CN	3.26/.51	3.29/.52
advanced	*RU*	2.01/.50	2.04/.46
	CN	3.02/.50	3.09/.42

4. Prosodic Parameters and Results of Analysis

The objective of analysis was to examine the relationships between German listeners' sentence-wise judgments on accentedness of the *CN* and *RU* data and objective prosodic speech parameters. First of all we counted the occurrences of some typical errors for each utterance:

Wrong accents: One type of error found in L2 utterances is the wrong placement of lexical stress in a word. Wrong choices of lexical stress position were only found in the CN data (e.g. "vor**sich**tig*"-"*cautious*"), but syllables inappropriately accented with respect to the sentence context were more frequent.

Additional pauses: Poor production performance results in additional pauses often triggered by difficult words and disrupting the speech fluency. These are not motivated by a deep prosodic boundary.

Additional syllables: As mentioned before, the insertion of epenthetic vowels creates additional syllables and therefore alters the rhythmic structure.

We looked at how the number of these events in an utterance was correlated with the judgments by the listeners. Results are shown in Table 3. As can be seen, wrong accent placements and additional pauses especially influence the prosodic judgment.

Table 3: *Correlations between the utterance-wise count of wrongly placed accents, additional pauses and additional syllables and the perceptual ratings.*

utterance-wise count of…	global rating	prosodic rating
wrong accents	.429	.530
additional pauses	.515	.607
additional syllables	.429	.426

Furthermore we examined phrase boundary tones at utterance-medial and final positions in questions and non-questions. Whereas all declaratives exhibited low final boundaries in sets *DE*, *RU* and *CN*, the situation differed in utterance-medial position as well as question-finally. In the case of y/n-questions Germans unanimously employ high boundary tones. Figures are 72% for Russian and only 46% for Chinese speakers. WH-questions by German speakers exhibit low boundary tones in 92% of cases, 83% for Russians and 59% for Chinese. There is a clear preference for high boundary tones utterance-medially for the Germans (81%), whereas the number is lower for Russian (56%) and Chinese speakers (54%).

For further quantitative analysis, the syllabic labels from the *DE, RU* and *CN* data were compared with respect to mean and standard deviations as well as rhythmic properties of the utterances. Analysis showed a considerably higher syllable rate of 5.7 syllables/second for the German speakers against 4.1 syllables/second for the Russian and 3.1 syllables/second for the Chinese learners of German. Within the *CN and RU* data, syllable rate in a single sentence is significantly negatively correlated with the global judgments of accent strength (ρ=-.663) and even more strongly with the prosodic judgment (ρ=-.750).

We then investigated whether the syllable-timed property of Russian and Chinese as opposed to the stress-timed properties of German also affected the realizations of the learners.

Based on the prosodic annotations of the realizations we categorized all syllables as either accented or unaccented. Results show that accented syllables in the *DE* corpus are relatively longer than unstressed syllables than in the *CN and RU* data. Table 4 shows the results of comparison.

Table 4: *Mean and standard deviation of syllabic durations for unstressed, potentially stressed and stressed syllables.*

set	syllable type	mean[ms]	s.d.[ms]	N
DE	accented	239	82	459
	unaccented	146	62	1005
RU	accented	299	95	291
	unaccented	205	91	659
CN	accented	342	135	584
	unaccented	254	115	1211

The ratio of mean durations unaccented/accented is .61 for the German speakers whereas it is .69 for the Russian and .74 for Chinese speakers. This suggests that although Russian and Chinese speakers apply the rules of the German accent system they tend to produce syllables of more uniform lengths than the German speakers.

Figure 3 shows averaged utterance-based means and standard deviations of PSLR calculations for the DE, RU and CN sets. The results confirm that the speech rate for the Russian and Chinese speakers is lower than for the Germans, but also shows that their rate variation is smaller, though comparable in proportion to their average speech rate.

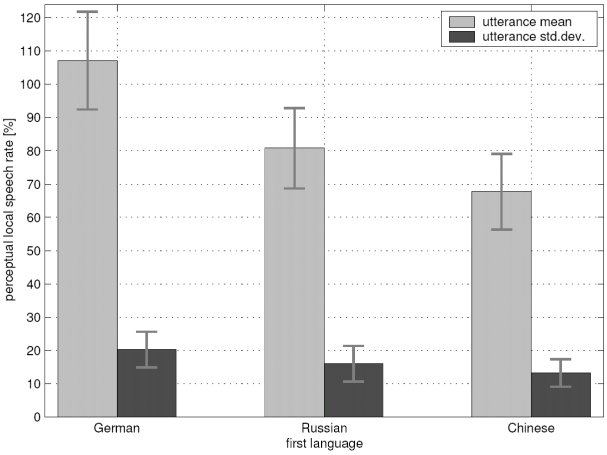

Figure 3: *Utterance-wise means and standard deviations of perceived local speech rate for German, Russian and Chinese speakers*

Looking more closely at the rhythmic patterns of individual sentences, we correlated the syllabic durations in one realization of a sentence with the syllabic durations in all the other realizations of the same sentence. The advantage of this approach is that the effect of the speech rate on this measure is rather small. This measure was previously used for evaluating the quality of a duration-predicting model in text-to-speech synthesis [11] and also in our earlier study on Australian English [4]. For easier comparison the duration of occasional short intra-utterance pauses is added to the duration of the syllable preceding that pause. This strategy can be justified by the fact that a pause in principle is an extreme case of the final lengthening usually observed in syllables preceding prosodic boundaries. Results indicate that the *DE* realizations (mean ρ=.897) are much more similar in their rhythmic structure (more highly correlated) than the *RU* (mean ρ=.737) and *CN* ones (mean ρ=.608).

In order to test whether the sentence-based correlations we had found were valid indicators of foreign accent we calculated the centroid of all *DE* utterances for each sentence. That is, for each syllable in a given sentence we averaged over all observed instances in the *DE* data set, yielding prototypical syllabic durations for each sentence. Subsequently we calculated the correlations between each of the *RU* and *CN* utterances and their corresponding *DE* duration norm. Statistical analysis showed that this rhythmic correlation was significantly (ρ=-.470) correlated with the German listeners' global judgment of foreign accent, however, slightly less with their prosodic judgments (ρ=-.455)

The extracted *F0* contours were modelled using the Fujisaki model in order to establish the differences between the *DE, RU* and *CN* data sets. To this effect automatic parameter extraction was performed on utterances of three of the sentences [12]. Then the analysis results were inspected and if necessary corrected using the interactive *FujiParaEditor* [13].

Table 5: *Mean and standard deviation of accent command amplitude Aa and accent command duration.*

set		*Aa*	duration [ms]
DE	mean	.29	294
	s.d.	.13	154
RU	mean	.38	239
	s.d.	.20	108
CN	mean	.31	243
	s.d.	.17	136

The numerical results of the analysis are displayed in Table 5. It shows means and standard deviations of accent command amplitude and duration for the DE, RU and CN data. As can be seen - though mean durations in sets RU and CN are quite similar - the Russians employ *F0* much more for marking accented syllables than the German and Chinese speakers. This is reflected by the higher values of accent command amplitudes *Aa*. In contrast, accent commands are longer for German speakers. This is due to the fact that groups of words are often connected with one and the same accent command (see example Figure 2, where the word group "gewusst haben willst" is associated with a single accent command).

If we look at the frequency of accent commands there are 1.54 commands per second in the *DE* group and 1.56 for the Russian and 1.63 in the Chinese group. The syllable-based frequency is one command every 3.73 syllables in the *DE* group but one command every 2.68 syllables in the *RU* and 2.14 in the *CN* data.

Table 6 shows means and standard deviations for the phrase command magnitude *Ap* which indicates the amount of *F0* reset taking place at the onset of a new phrase.

Table 6: *Mean and standard deviation of phrase command magnitude Ap.*

set	mean	s.d.
DE	.29	.15
RU	.39	.16
CN	.27	.12

It is obvious that the Russian speakers adjust their declination line more strongly which is an indication that they employ a larger F0 range when they talk than the German and Chinese speakers. They also rephrase more frequently, on the average once every 5.88 syllables compared to 8.33 syllables for the German speakers. The Chinese speakers rephrase even more often, namely every 5.26 syllables. This result however might also be partly due to the higher speech rate of the Germans.

5. Discussion and Conclusions

The current study concerned the prosodic analysis of accented German speech data produced by Russian and Chinese learners. We found that the number of additional pauses in an utterance is strongly correlated with the percept of foreign accent. The same holds for the number of wrongly accented syllables, as well as the number of additional syllables due to the insertion of epenthetic vowels. The latter phenomenon is typical of the Chinese learners. In contrast to an earlier study [5] we also found a strong correlation between the average speech rate in an utterance and the strength of perceived accent.

On the rhythmic level Russian and Chinese learners of German produce relatively longer unaccented syllables than German speakers, which suggests, that their rhythm is influenced by the syllable-timed structures of Russian and Chinese. The syllabic durations in the German group are more uniform than those within the Russian and Chinese group expressed by the durational correlations between individual productions of the same sentence. On the intonation level Russian speakers produce stronger excursions of F0 and use a wider range of F0 than the German controls. Russians and Chinese place pitch-accents more frequently than their German counterparts and do not connect prosodic word groups as much. Accent commands are therefore shorter and clearly connected with the individual syllables. At intermediate prosodic boundaries, Russians and Chinese show less preference for high boundary tones than the Germans.

Future work will concern perceptual experiments with segmentally and prosodically manipulated stimuli in order to examine which factors contribute most to the percepts of strong foreign accent and reduced intelligibility. Furthermore, we will test whether our findings can be applied to enhance computer-aided pronunciation training, especially the apparent rhythmic and melodic differences observed.

6. References

[1] Anderson-Hsieh, J., Johnson, R. and Koehler, K., "The relationship between native speaker judgements of nonnative pronunciation and deviance in segmentals, prosody and syllable structure", Language Learning 42(4), 529-555, 1992.

[2] Magen, H.S., "The perception of foreign-accented speech", Journal of Phonetics, 26(4), 381-400, 1998.

[3] Bissiri, M.P. and Pfitzinger, H.R., "Italian speakers learn lexical stress of German morphologically complex words", Speech Communication 51(10), 933-947, 2009.

[4] Nguyen, T. and Ingram, J., "A corpus-based analysis of transfer effects and connected speech processes in Vietnamese English", Proceedings of the Tenth Australian International Conference on Speech Science & Technology, Sydney, Australia, 2004.

[5] Mixdorff, H. and Ingram, J., "Prosodic Analysis of Foreign-Accented English", Proceedings of Interspeech 2009, Brighton, England, 2009.

[6] Jokisch, O., Wagner, A. et al., "Multilingual Speech Data Collection for the Assessment of Pronunciation and Prosody in a Language Learning System", Proceedings of Specom 2009, St Petersburg, Russia, 2009.

[7] http://www.wevosys.com.

[8] http://www.praat.org.

[9] Fujisaki, H. and Hirose, K., "Analysis of voice fundamental frequency contours for declarative sentences of Japanese", Journal of the Acoustical Society of Japan (E) 5(4) 233-241, 1984.

[10] Pfitzinger, H.R., "Local Speech Rate Perception in German Speech", Proc. ICPhS 1999, 893-896, 1999.

[11] Mixdorff, H. and Jokisch, O., "Evaluating the quality of an integrated model of German prosody", International Journal of Speech Technology 6(1): 45-55, 2003.

[12]Mixdorff, H. "A novel approach to the fully automatic extraction of Fujisaki model parameters", Proceedings of ICASSP 2000, vol. 3, 1281-1284, Istanbul Turkey, 2000.

[13]http://public.beuth- hochschule.de/~mixdorff/thesis/fujisaki.html.