German Learners of Japanese - Perceptual and Prosodic Analysis of Utterances from a Meditative Setting

Hansjörg Mixdorff[1], *Ryoko Hayashi*[2], *Yoriko Yamada-Bochynek*[3], *Keikichi Hirose*[4], *Hiroya Fujisaki*[5]

[1] Dept. of Computer Science and Media, Beuth University of Applied Sciences, Berlin, Germany
[2] School of Intercultural Studies, University of Kobe, Japan
[3] Institute of Japanese Studies, Free University Berlin, Germany
[4]Department of Information and Communication Engineering, the University of Tokyo, Japan
[5] Professor Emeritus, the University of Tokyo, Japan

mixdorff@beuth-hochschule.de, rhayashi@kobe-u.ac.jp, yayo@zedat.fu-berlin.de, hirose@gavo.t.u-tokyo.ac.jp, fujisaki@alum.mit.edu

Abstract

This study examines how closely Japanese utterances produced by German learners match those of Japanese natives, especially with respect to word and sentence prosody. Utterances were rated perceptually by Japanese native listeners as well as evaluated with respect to timing and *F0* contours. We found that the German learners speak at a similar speech rate, but their syllabic durations are much more at variance than those of the Japanese controls. Altogether their rhythmic patterns are more similar than those by the Japanese. Contrary to our expectations perceptually prominent so-called pseudo-accent syllables are not lengthened, although the timing of tonal transitions at these is slightly different from the Japanese norm. Some transitions were not at all realized due to "de-accenting" phenomena.

Index Terms: foreign accent, prosody, quantitative analysis

1. Introduction

This study examines Japanese utterances produced by German learners in a meditative setting supposed to calm them and help them focus on the upcoming lesson. This so-called "relax" setting is part of an unconventional teaching method called JaFIX developed by Yamada-Bochynek. We wanted to see how close these frequently and regularly uttered sentences come to those of Japanese natives, especially with respect to prosody. Since the learners were very much familiar with the sentences we expected to find something like an optimal case or a top-line of what degree of prosodic nativeness learners at an early stage of their training are able to attain. It is also a preparatory study laying ground for comparisons with more conventional language teaching methods.

For the sake of comparison we collected the same data from ten speakers of the Tokyo dialect. In addition, the German utterances were rated perceptually by Japanese native listeners with respect to their general quality, but also with respect to their prosody. Hence, the current study is centered on those deviations which are caused by inter-language interference between L1 to L2 and are common to a group of subjects with the same native language. According to Schulte-Pelkum, this kind of interference is the result of contrasts between L1 and L2 [1]. With respect to the phoneme inventory, Japanese is regarded as relatively easily acquired by Germans. In the case of vowels Neyer [2] reports 19 vowel phonemes for German and only 5 for Japanese all of which have similar, but not identical counterparts in German. The opposition of long and short vowels exists in Japanese as well as in German, although long and short vowels in German also exhibit quality differences which are not the case in Japanese. According to Neyer, German features 21 consonant phonemes against 13 in Japanese. All Japanese consonants find parallels in the German sound system. The Japanese syllabic system is also comparably simple and few errors by German learners should be expected due to the syllabic structure.

The dominant acoustic feature of the lexical accent in Japanese is the F0 contour. Each mora of a Japanese word is assigned either a high or low tone. A transition from low to high or from high to low occurs after the first mora. In the former case a high-low transition may be found after the Mth mora. Hence theoretically there are N+1 possible accent patterns for N-mora words. The type without transition from high to low is called flat or unaccented type and denoted (N,0) whereas the others are denoted (N,M). For ease of representation Fujisaki's convention will be followed which denotes the accented word type by 'D' for 'downfall' and the unaccented type by 'F' for 'flat'[3].

In Japanese, the accent type hardly affects the syllable durations as was found in a study on similar two-syllable words of the two languages [4]. In contrast, in German accented syllables are usually longer than unaccented ones. The study also showed that the accentuation of Japanese words in terms of on- and offsets of accent commands is subject to less timing variation than in German, indicating its close temporal relationship with a unit of time (i.e. the mora) typically shorter than a German syllable. In the current study we are therefore interested in the question whether German learners unconsciously transfer their accentuation habits to Japanese, that is, whether, for instance, they lengthen morae that are connected to F0 rises or falls and which therefore might be perceived as more prominent. It is examined how the German learners, whose L1 lexical accent system interacts strongly with the sentence prosody, master the seemingly similar prosodic surface structures of Japanese which are largely determined by the accent types of the words involved. Japanese D-type words, for instance, exhibit falls of F0 which resemble declarative sentence intonation in German whereas F-type words tend to end on a medium level of f0 which in turn resembles non-terminal sentence intonation in German.

Furthermore we examine the alignment of rising and falling tone switches by Japanese and German subjects.

2. Speech Material

The material examined in the current study was collected during lessons of Japanese following the ***JaFIX*** method developed by Yamada-Bochynek[5]. JaFIX is based on "Evolutionary Cultural Semiotics" advocated by Walter A. Koch [6], encompassing a whole range of developmental and epistemological features of onto- and phylogenesis of human speech and cognition. Classroom activities take place on the "JaFIX carpet" and desks and chairs, two of the major "shackles" in a traditional setting of school culture, are put aside, providing the class an arena for *corporeal* learning and various kinds of multimodal exercises. Ten German participants of the Japanese Basic Course I at Japanese-German Center Berlin (8 male, 2 female) were recorded during the 20 minute "relax" phase preceding every 2 hour lesson, after an 8 week long learning period of 2 hours twice weekly, a total of 32 hours. Students had been provided with the relax materials as well as recordings to practice individually. The relax sentences are introduced in both languages at the beginning, but in the course of the next five lessons, however, the German semantics become less and less, and by the sixth time, the relax is performed solely in Japanese. From the 6th lesson on the "acting out" begins – the corporeal implementation of the text's content with fluent bodily movements, mirroring the features of the text. The class gets up when the induction is finished and follows the corporeal movement of the inducer, along which the text will be uttered. The acting-out is followed by shadowing. After the 9th week the participants start to take over the role of induction, leading the acting-out and shadowing. The recording begins with the first student who utters the sentences from his/her memory.

Figure 1: *Students during the relax phase of the JaFIX lesson.*

In order to compare the utterances by the German learners (ultimately a total of 57 utterances were admitted to analysis), ten native speakers of the Tokyo dialect (5 male, 5 female) were recorded. They were instructed to read the sentences slowly and relaxed, yielding a total of 60 utterances.

3. Method of Analysis

All data were forced-aligned at the University of Tokyo. The resulting phone segmentations were manually checked and converted into a mora tier. *F0* values were extracted at a step of 10ms using the *PRAAT* default pitch extraction settings [7]. Then Fujisaki model [8] parameters were estimated [9] (alpha=3/s, beta=20/s) and manually corrected in the *FujiParaEditor*[10], taking into account the accent types of the underlying text in the Japanese cases.

Table 1: *"Relax" sentences used in this study. The table displays the romanized Japanese text and its English translation. High tone morae are indicated by grey text background, arrows indicate critical tonal transitions at mora boundaries. Pseudoaccent syllables (see text) are underlined.*

1)	Shi↓ zuka ni o↑chitsuite yu↑Q ta↓ri to shiteiru *I feel quiet, relaxed and balanced.*
2)	a↑ shi↓ga yu↑kani tsu↓iteiru no o ka↑Njiru *I feel my feet attached to the floor.*
3)	mi↑giteto mi↑giudega yu↑kani tsu↓iteite, mi↑giteto mi↑giudeno o↑mosa o kaNjiru *The right hand and right arm are attached to the floor, I feel the weight of the right hand and arm.*
4)	wa↑tashi no (↓) mi↑giteto ↑mi-/↑giudewa (↓) o↑mo ↓i. o↑mo ↓kute ki↑mochi ii *My right hand and right arm are heavy, pleasantly heavy.*
5)	Ka↑rada zeNtai ga (↓) yu↑ka ni tsu↓iteite, ka↑rada zeNtai no (↓) o↑mosa o kaNjiru *The whole body is attached to the floor, I feel the weight of the whole body.*
6)	Wa↑tashi no (↓) ko↑kyu: no ri↓zumu wa↑tashino (↓) ko↑kyu: no ugoki *The rhythm of my breath, the movement of my breath.*

In Table 1, high tone morae are indicated by grey text background and low ones by white. Arrows indicate the position and direction of tonal transitions at mora boundaries. At this stage we are not so much interested in the identity and exact boundaries of phones actually realized, but the rhythmic structure of the utterances. As a working hypothesis we assume that German listeners perceive tone switches occurring in Japanese words as prominence-lending features, since in German they are associated with accented syllables. This assumption possibly combined with other factors such as structural similarities with German words or the structure of the syllable itself might lead to an interpretation of certain Japanese morae as being more salient, attaining what we will call in the scope of this paper the status of "pseudoaccent syllables". Since a systematic way of identifying pseudo-accent syllables does not exist, the first author of this study who is fluent in Japanese marked up all the syllables in the text that he perceived as most prominent in the recordings of the German subjects. These are indicated in Table 1 by underlining (NB: in some cases a syllable comprises two morae, in others it is identical with the mora). Figure 2 and Figure 3 display examples of analysis of the initial part of sentence 5, spoken by a German and a Japanese male subject, respectively. Each panel contains from the top to the bottom: The speech wave form, the *F0* contour (+signs: extracted, solid line: Fujisaki model-based), the moraic SAMPA transcription, the underlying phrase and accent commands.

4. Perceptual Rating

A total number of 10 native speakers of Japanese (3 male, 7 female), most of them were graduate school students majored in linguistics or foreign language teaching at Kobe University, took part in a perceptual rating test. They were given questionnaires in the form of a Winword document containing hyperlinks to the sound files. We included all data of German speakers (57 sentences) and 12 sentences uttered by Japanese speakers as a reference in random order. 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 5 being native-like, 1 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. The rating experiment was done individually in a silent room using headphones. Table 1 shows the mean and standard deviation for the German learners and the Japanese speakers. The best German learner reached a mean of 3.67, the poorest 2.05 in the overall judgments. As can be seen, the prosodic ratings are slightly better than the overall ones.

Table 2: *Mean and standard deviation of perceptual ratings for Japanese native and German learners.*

speakers	global rating	prosodic rating
German	3.02 /.45	3.24 /.12
Japanese	4.92 /.09	4.94 /.10

We found that inter-rater correlation of judgments was above .7 in all cases. Split correlation between two randomly selected groups of five raters each is .943 and .869 for overall and prosodic judgments, respectively. The correlation between overall and prosodic judgments is .921.

5. Prosodic Parameters and Results of Analysis

Timing Analysis. For further quantitative analysis, the moraic labels from the Japanese and German data were compared with respect to mean and standard deviations as well as rhythmic properties of the utterances. Analysis showed that the mora rate of the German speakers (5.84/s) is only slightly lower than that of the Japanese (6.04/s).

We found that within the German data, mora rate in a single sentence is not correlated with the global judgments of accent strength or prosodic performance. We then investigated whether the mora-timed property of Japanese as opposed to the stress-timed properties of German also affected the realizations of the learners.

We therefore compared the durations of pseudoaccent syllables in the two groups which we expected to be relatively longer for the Germans. Table 3 shows that this is not the case. However, standard deviations are generally larger in the German group.

Table 3: *Mean and standard deviation of moraic durations for pseudoaccent morae and others.*

set	mora type	mean[ms]	s.d.[ms]	N
German	pseudo acc.	172	61	324
	Others	171	81	1058
Japanese	pseudo acc.	167	43	349
	Others	165	63	1131

This is also reflected by the larger mean normalized pair-wise inter-variability index *npvi* [11] of 63.69 for Germans as compared with 49.85 for the Japanese subjects.

Looking more closely at the rhythmic patterns of individual sentences we correlated the moraic durations in one realization of a sentence with the moraic 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. It was previously used for evaluating the quality of a duration-predicting model in text-to-speech synthesis and also in an earlier study on Australian English [12] spoken by Vietnamese learners. Results indicate that the German realizations (mean ρ=.722) are more similar in their rhythmic structure (more highly correlated) than the Japanese ones (mean ρ=.670). The inter-group correlation is .628. This result seems surprising, since it contradicts findings from [12] where the native controls were usually more highly correlated.

If we consider, however, that the Germans had been exposed to the same Japanese reference for a long time and that the Japanese were recorded individually and without specific acoustic target, this result seems plausible. In order to test whether the sentence-based correlations we had found were valid indicators of foreign accent we calculated the centroid of all Japanese utterances for each sentence. That is, for each syllable in a given sentence we averaged over all observed instances in the Japanese data set, yielding prototypical moraic durations for each sentence. Subsequently we calculated the correlations between each of the German utterances and their corresponding Japanese duration norm. Statistical analysis showed that this parameter is significantly (ρ=.402, $p < .01$) correlated with the Japanese listeners' judgment of prosodic performance, however, less so with the general judgments of foreign accent strength (ρ=.278, $p < .05$).

Intonation Analysis. We analyzed the intonational characteristics of Japanese and German speakers by comparing the Fujisaki model parameters for each group. Mean and standard deviations of accent command amplitudes are displayed in Table 4. As can be seen Japanese speakers use a wider *F0* range than Germans, reflected by the higher values of accent command amplitudes *Aa*. Table 5 shows means and standard deviations for phrase command magnitude *Ap* which indicates the amount of *F0* reset taking place at the onset of a new phrase. At least the German males rephrase more weakly than the Japanese subjects whereas the two German females use *F0* resets of the same strength as the Japanese.

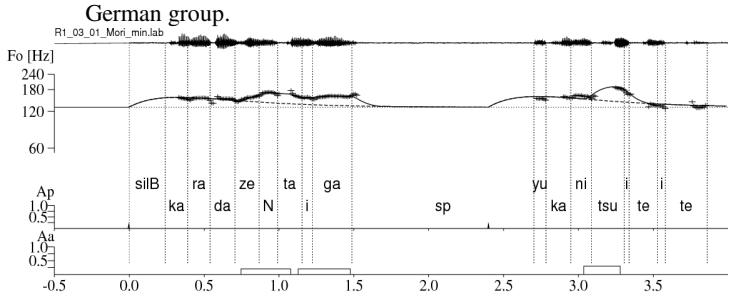

Figure 2: *Example of analysis, first part of sentence (5) "Karada zentai ga yuka ni tsuite ite" "-" The whole body is attached to the floor." uttered by a male German.*

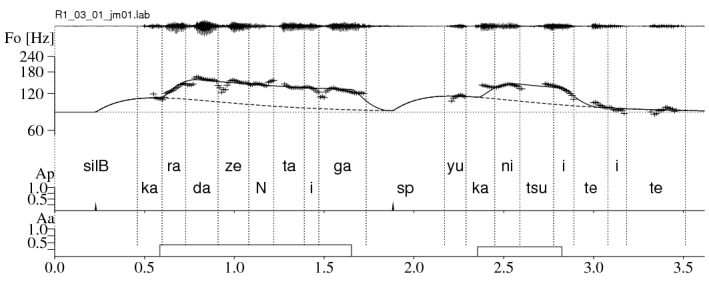

Figure 3: *Example of analysis of the same sentence as in Figure 1, uttered by a male Japanese. The f0 rise after 'ka' is missing in the German example where it instead occurs after the mora 'zeN', the syllable erroneously accented by the learner (see text).*

Table 4:*Means and standard deviations of accent command amplitudes Aa, number of accent commands for Japanese and German speakers.*

set	gender	mean	s.d.	N
Japanese	male	.30	.12	124
	female	.29	.12	120
German	male	.24	.12	188
	female	.24	.12	50

Table 5: *Means and standard deviations of phrase command magnitude Ap, number of phrase commands.*

set	gender	mean	s.d.	N
Japanese	male	.20	.08	67
	female	.20	.11	64
German	male	.15	.09	99
	female	.20	.08	34

We examined the alignment of tone switches in the vicinity of the pseuoaccent morae in order to see whether Germans and Japanese behaved differently. The timing is calculated as the distance between the moraic boundary where the tonal transition is supposed to occur in the Japanese word and the accent command onset time T1 in the case of rising *F0*, and the distance between the moraic boundary and the accent command offset time T2 in the case of falling *F0.* This means that a negative value indicates that the tonal transition occurs before the boundary, and a positive value indicates that the transition takes place after the boundary. In Figure 3, for instance, the first accent command begins -24 ms from the offset of the mora [ka], marking the beginning of the rising f0 movement. The result of analysis is displayed in Table 6. The figures suggest that rising movements start slightly earlier in the Japanese subjects than in the Germans and are more closely timed with the mora boundaries as indicated by the smaller standard deviation. In contrast, falling movements show similar standard deviations in both groups and occur slightly earlier in the German subjects than in the Japanese. None of these differences, however, are statistically significant. If we examine more closely those pseudoaccent syllables associated with a rising tone switch we find that for some words the Japanese word accent pattern prescribes a tonal transition at the beginning of the respective mora, yielding a high tonal target as in wa↑tashi, for instance, and a tonal transition at the end of the respective mora for other words, such as in yu↑ka where the next mora attains a high target. If the Germans were to produce a rising tone switch in order to lend the pseudo accent syllable more prominence, we would expect the timing of the tonal transition to be the same for both types in the German subjects whereas the Japanese should show a clear distinction between early and late rises. To test this hypothesis we calculated the distance between T1 and the onset of the respective mora. Although there is a clear difference between early and late types in the Japanese (48 ms vs. 112 ms) there is also a smaller difference in the German subjects (68ms vs. 101ms). Furthermore, not all required transitions were actually produced by the German learners. If we compare Figure 2 and Figure 3, it becomes obvious that the German speaker fails to realize the *F0* rises on the morae [ka] and [yu] and instead produces these on the morae [ze] and [tsu] which are also perceptually most prominent. This is a typical example of "de-accenting" due to L1 prosodic habits.

6. Discussion and Conclusions

The current study concerned the prosodic analysis of Japanese speech data produced by German learners. We found that despite comparable speech rate, the German learners are more at variance with respect to syllabic duration though they do not exhibit lengthening of prominent syllables. They are also more uniform with respect to their duration patterns, possibly due to having shared the same Japanese reference. Tonal transitions resemble those produced by Japanese subjects though some are deleted due to "de-accenting". However, we have to bear in mind that the number of tokens examined is too small for reaching general conclusions. Future work will concern comparisons with learners having been exposed to more conventional teaching methods and employing different elicitation strategies.

Table 6: *Alignment of rising and falling F0 movements with respect to mora boundaries, at which tonal transitions occur in ms.*

set		rising	falling
German	mean	12	10
	s.d.	125	113
	N	199	84
Japanese	mean	-5	21
	s.d.	75	112
	N	228	89

7. Acknowledgements

Thanks go to Ryusuke Mihara at the University of Tokyo for performing *F0* extraction and forced alignment.

8. References

[1] Schulte-Pelkum, R., Interferenzfehler bei deutschsprechenden Japanern. In G. Stickel, editor, Deutsch-japanische Kontraste, pp. 59-111. TBL Verlag Gunter Narr, Tübingen, 1976.

[2] Neyer, F.A. Die segmentalen Phoneme des Deutschen und Japanischen in kontrastiver Betrachtung. Beiträge zur deutschen Literatur, pp. 1-28, 1971.

[3] H. Fujisaki, K. Hirose, and N. Takahashi. Manifestation of linguistic and paralinguistic information in the voice fundamental frequency contours of spoken Japanese. In Proceedings of 1990 ICSLP 1990, pp. 485-488, Kobe, 1990.

[4] Mixdorff, H. Fujisaki, H. Word accent in German and Japanese - a comparative study using a quantitative model. In Reports of the 1995 Spring Meeting of the Acoustical Society of Japan, vol. 1, pages 313--314, 1995

[5] Yamada-Bochynek, Y.(2009), „KK KanjiKreativ: Ein e-Learning-Programm zum Erlernen der 1945 Jōyōkanji…", in Árokay, J., et al.*Essays in Honour of Irmela Hijiya-Kirschnereit.* München: Iudicium, 501-523.

[6] Koch, W. A. (1986). *Evolutionary Cultural Semiotics*. Bochum: Brockmeyer; (ed.) (1989), *Culture and Semiotics*: Bochum: Brockmeyer;.

[7] www.praat.org.

[8] 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.

[9] 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.

[10] http://public.bht-berlin.de/~mixdorff/thesis/fujisaki.html

[11] White, L., & Mattys, S.L., "Calibrating rhythm: First language and second language studies", Journal of Phonetics, 35, 501-522, 2007.

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