Native-like Duration Ratio of Stressed vs. Unstressed Syllables through Visualizing Prosody

Markus Rude

Institute for Liberal Arts and Sciences, Nagoya University, Japan

mrude@ilas.nagoya-u.ac.jp

Abstract

Japanese language learners of stress-timed languages like English (L2) or German (L3) have problems acquiring the prosody of these languages. The questions are, how to teach prosody in this context efficiently, and – specifically – what role visualizations play in the acquisition of accentuation. Therefore a reading experiment was undertaken, as reported here, in which students read out texts in three formats, among them a curved text with prosody-encoding size variations called prosodic writing (PW). In a pre-test, students read a normal text; in a post-test they read PW. Additionally, access possibilities to corresponding audio files varied among students. The following results were obtained for the accentuation of the German word "zentral" [tsɛnˈtraːl]: (1) there was a substantial increase in correct accentuation among the 15 students reading PW; (2) of these, all five students without access to audio files corrected their accentuation; and (3) for four of these five, the duration ratio of stressed to unstressed syllables approached the ratio of a native speaker (>2:1). Visualizations might therefore be an efficient tool for teaching certain aspects of prosody.

Index Terms: speech prosody, prosodic writing (PW), interference, visualization, stressed syllable, duration ratio.

1. Introduction

It has been reported that Japanese language learners of stress-timed languages such as English or German are able to acquire stress patterns similar to those of English or German native speakers (NS). However, there are two major differences: (1) Japanese language learners mainly use the repertoire of the Japanese language system and thus tend to perceive and realize stress by falling pitch accents rather than by a combination of (rising or falling) pitch accents, duration, intensity and tension, and (2) though these non native speakers (NNS) will eventually start to use the mentioned stress correlates, they will do so much less than NSs do in quantitative terms [1].

These two differences have consequences for both, perception and production: (1) sometimes there might be a difficulty in recognizing words as being stressed if they are marked by a rising pitch accent, and (2) there might be a problem in developing a sense of the rhythm of the language, since rhythmic phenomena depend largely on temporal relationships of adjacent syllables, e.g. time intervals between stressed syllables, and duration ratios between stressed and unstressed syllables.

The broad research question is how to teach prosody with all its phenomena in an efficient way, in particular in cases of language interference between languages. The narrow research questions are: What impact do visualizations of prosody have on the acquisition of accentuation and its prosodic correlates? Can such visualizations offer advantages compared to audio-based prosodic input.

2. A reading test as an experiment

In order to discover pronunciation problems in the suprasegmental domain, a pair of reading tests – a pre-test and a post-test – were recorded and analyzed.

2.1. The procedure of the reading test

There were three classes, A, B, and C, each of which had about 30 students, mostly Japanese and some Chinese. The subgroup of 15 students relevant to this paper were students with Japanese as L1, and – except one who had lived in the U.S.A.– none of them had been abroad for more than three months. Every student read aloud one out of 5 short texts in front of the class. (Thus, each text was read about 18 times, and in total, there were about 90 read-out texts)

- In a *pre-test*, the students from A and C read these 5 texts from their common textbook, whereas the students from B received copies from that textbook page. A & C students were both familiar with the written texts, which were covered in class some weeks before, partially also with their prosodic realizations, which were on the textbook's audio CD.
- Class A had about 10 minutes to practice before reading. Class B had the same time for reading practice, but an additional 20 minutes for text comprehension. In contrast, class C had received a one-week preparation period for the pre-test.
- After the pre-test, a *short course* on prosody was given: the basic components of prosody were explained (intonation, stress, rhythm, pausing) and introduced as part of five evaluation criteria (loudness, clarity, intonation or non-flatness, stress, and fluency as speed/smoothness) for the post-test.
- This *short course* included also practice time for reading aloud an example in underlined form (long/short underlines showed primary/secondary stress) and in prosodic writing (PW, see section 2.4). The training consisted of listening to the example while reading along with the marked text, listening and repeating, and finally reading in unison while trying to realize the same stress patterns or intonation contours as shown in the text.
- In addition, the students received the 5 texts from the pre-test in the new formats: two with underlined stressed syllables, two in PW, and one in normal writing for self marking of stress. An announcement was made that students could choose freely which format they would read in the post-test in the following lesson.
- In the *post-test*, every student read the same text as in the pre-test.

Both, pre-test and post-test were recorded with an MP3-/WAV-Recorder (Roland Edirol R-09HR).

2.2. The analysis of the audio recordings

In a first step, the software f5 by Audiotranskription.de [2] was used to transcribe portions of the audio recording and to link the roughly 800 lines of transcription to the audio file via the same number of time stamps.

In a second step, the time-stamped lines were re-ordered: corresponding student productions from pre-test and post-test were positioned next to each other. At this stage, a perceptual analysis was performed and phenomena of interest were selected for further analysis. Only qualitative decisions were made, e.g. on stress location. Some of these phenomena were also reported at [3].

In a third step, the productions related to these phenomena were analyzed in a quantitative way through Praat [4]; the results will be presented in the following sections.

2.3. The nuclear accent of a NS

One very obvious result could be obtained from the following text:

"Ich wohne in einer Altbauwohnung. Meine Wohnung ist groß und hell, und sie liegt zentral. ..." [5].
[I live in an apartment in an old building. My apartment is spacious and bright and it is located in the center. ...]

The male NS from the textbook CD clearly stressed the very last syllable of the utterance ("tral" in "zentral"). The prosodic correlates duration, frequency, and intensity can be seen in Table 1: the duration of the voiced sections of both syllables, the number of glottal oscillations (F_0-wave cycles), the average values of the fundamental frequency F_0 and intensity in the voiced sections; Figure 1 shows the same data in a chart: in the upper part the corresponding waveform (black) and the glottal pulses (blue verticals), in the lower part the intensity curve (green, continuous) and F_0 curve (blue dotted, interrupted).

Table 1. *Parameters of the NS's production of "zentral" [tsɛnˈtra:l]. Stress can clearly be perceived on the 2nd syllable "tral" and is reflected in the ratio of durations of the voiced syllable sections. (All values are rounded)*

NS's "zentral" [tsɛnˈtra:l]	Duration [ms]	# of waves	F_0 [Hz]	Intensity [dB]
"(z)en(t)" [tsɛnt]	110.1	12	109.0	79.7
"(t)ral" [tra:l]	271.6	28	103.1	79.4
ratio (ral/en) [%]	**247%**	**233%**	**95%**	**100%**

The values in Table 1 reveal that duration is the main physical correlate of stress. It seems as if F_0 and intensity do not reflect stress at all (with almost identical values for both syllables). However, due to the phenomenon of declination [6], F_0 decreases towards the end of intonational phrases (IPs), and thus the almost same F_0 on "(t)ral" compared to "(z)en(t)" (just 5% lower) is still perceived as a pitch accent, higher than the second last syllable. For the same reason, loudness of the last syllable will also be perceived as greater than that of the second last, though acoustically it is slightly smaller in intensity.

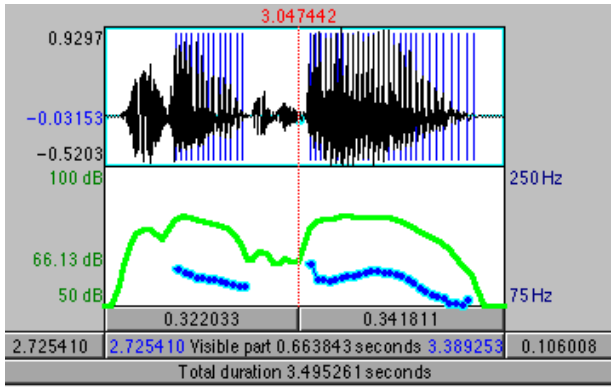

Figure 1: *Waveform (black), glottal pulses (blue), intensity curve (green) and F_0 frequency curve (blue) of the NS's production of "zentral" [tsɛnˈtra:l]. The duration of the voiced segments in the 2nd syllable, "(t)ral" (right), is more than twice as long as that of the voiced segments in the 1st syllable "(z)en(t)" (left). (Fig. 1 and Fig. 3 were made with Praat [4][1])*

2.4. The concept of prosodic writing (PW)

As for the text example from the last section, the visualization for the post-test was in the format of prosodic writing (PW). PW is similar to the usual text in following the orthographic standards. However, PW additionally expresses prosodic features through defined distortions, as if the text would be curving in 3D-space. PW has been introduced for example in [7].

The visualization of prosody in PW is straightforward. It captures durational phenomena in the horizontal, 1st dimension; however, it visualizes psychological time rather than physical time. Furthermore, it captures the pitch contour through the up-and-down curvature in its vertical, 2nd dimension, and loudness through letter size or boldness in its "depth", pseudo-3rd dimension.

Readers familiar with Bolinger's "Intonation" will remember in his introduction his way of visualizing the pitch contour by placing the letters of a written sentence on different height [8]. PW can be seen as an extension of this approach by "stretching" graphemes in the horizontal to express duration, and "blowing up" graphemes to express loudness while maintaining the integrity of the visual appearance of words.

The word "zentral" should therefore be visualized through a long-stretched second syllable, because this syllable carries primary stress. Since PW visualizes perceptive features (pitch and loudness) rather than acoustic features (F_0 and intensity), the second syllable may be written higher and larger compared to the first syllable: it sounds higher and louder subjectively, though it is neither higher in F_0 nor more intense objectively. Fig. 2a shows a possible visualization of "zentral", Fig. 2b the corresponding context.

[1] Version: 4.4.11, setting for analysis in tables: Standard setting for both, pitch (50 Hz - 500 Hz, optimized for intonation (AC method)) and intensity (50 dB – 100 dB; averaging method: mean energy; subtract mean pressure). Same setting for figures except for pitch (75 Hz – 250 Hz) for visual reasons (contours become higher).

Figure 2a: *Visualization of nuclear accent on the 2nd syllable in "zentral" [tsɛnˈtra:l] in PW. The graphemes of "ral" extend more than twice as long in the horizontal (the time dimension), compared to the graphemes of "en"; the two grapheme sequences represent the voiced segments of the syllables "tral" and "zent", respectively.*

Figure 2b: *The stressed word in context.*

2.5. Results from the reading test of the NNSs

This text was read by 15 students as normal text in the pre-test, and one week later in PW in the post-test (except the 4 students from class C, who – due to a holiday – had two weeks between the pre-test and the post-test). In the following, the main results from the reading experiment are summarized:

- In the pre-test, only one student (class A) out of the 15 stressed the 2nd syllable correctly (6.7%). All students from B and C stressed the 1st syllable.
- In the post-test, 11 of 15 students stressed the 2nd syllable correctly (73,3%). From class B, all 5 students corrected their accentuation. This is remarkable, since this class did not use any audio files.
- The numerical results were similar among the 5 students from B. The ratio of durations was modified from about 1:1 to more than 2:1 in all measurable cases. (265%, 241%, ?, 212%, 220%)[1].
- The modification in the post-test did not only involve lengthening of the stressed syllable, but also shortening of the unstressed syllable to 80%, in one case to almost 50% of the duration from the pre-test. (78%, 51%)[2]

[1] In the post-test, one student's voice ended with a creaky voice, so that the duration of the intonation contour could not be measured as in the case of the other four students.

[2] In the pre-test, some students had produced a voiced plosive [d] instead of the unvoiced plosive [t]; the resulting continuous intonation contour could not be divided and measured like the other contours, thus the post-test/pre-test ratio could also not be calculated.

2.6. The numerical data of one NNS as an example

In the following, the data of one student from pre-test and post-test are being shown. Table 2 shows the numerical data of the student, Fig. 3 the corresponding graph.

Table 2. *Parameters of NNS's production of "zentral". Mean values of F_0 and intensity in voiced sections. In pre-test (upper half), stress is perceived on 1st syllable "zen" (wrong). In post-test (lower half), stress is perceived on 2nd syllable "tral" (correct).*

NNS B26 pre-test	Duration [ms]	# of waves	F_0 [Hz]	Intensity [dB]
"(z)en(t)" [tsɛnt]	146.5	17	116.0	74.1
"(t)ral" [tra:l]	145.9	14	96.0	75.5
ratio (ral/en) [%]	**100%**	**82%**	**83%**	**102%**
NNS B26 post-test	Duration [ms]	# of waves	F_0 [Hz]	Intensity [dB]
"(z)en(t)" [tsɛnt]	114.4	14	122.4	76
"(t)ral" [tra:l]	251.6	29	115.3	76.6
ratio (ral/en) [%]	**220%**	**207%**	**94%**	**101%**

Fig. 3 shows the data of the same student from the pre-test (upper chart) and the post-test (lower chart): Waveform (black), glottal pulses (blue), intensity curve (green) and frequency curve (blue). In the pre-test, the voiced portions of both syllables (blue regions) are of similar duration, showing the tendency of Japanese speakers to produce syllables of equal length, instead of reducing unstressed syllables.

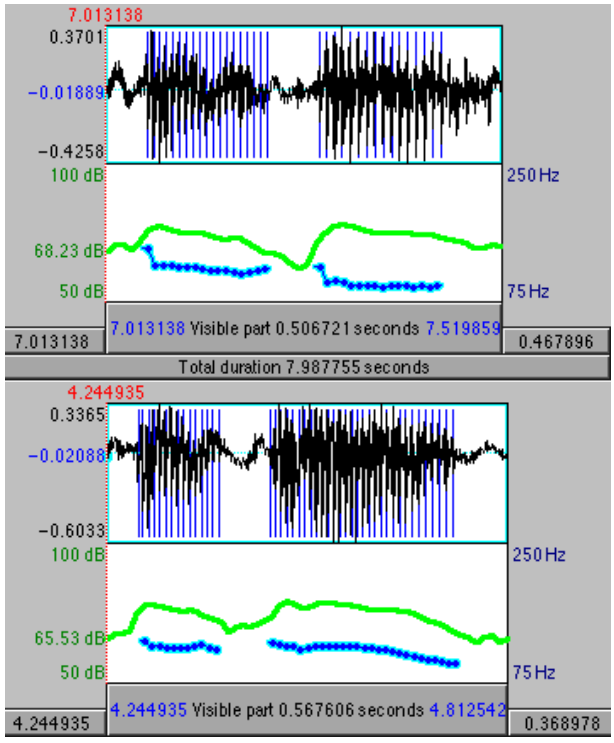

Figure 3: *Waveforms, glottal pulses, intensity and F_0 curves of the NNS (B26) of "zentral" in pre-test reading normal text (upper chart) and post-test reading PW (lower chart). The lower F_0 frequency curve (blue) is more similar to that of the NS (Fig. 1).*

In the post-test, the duration of the voiced portion "ral" (right) is more than twice as that of "en" (left), a similar ratio as for the NS from Fig. 1.

A comparison of the values from Table 2 and Table 1 confirms this result – the closeness of NS and NNS in the post-test – numerically, and an auditory comparison supports this finding perceptually.

3. Discussion

Some interesting results could be found when comparing the pre-test, in which students read aloud short texts, and the post-test, for which the students received the same texts in PW (prosodic writing, section 2.4). [1]

One result is the weak performance of class C (section 2.5). This class had one week preparation time for the pre-test, during which 3 of 4 students used the audio CD for preparation, but none of the 4 produced the correct stress pattern on "zentral".

Of course, there are two difficulties related to stress location on this word: most German words have stress on the first syllable, at least among the vocabulary for 1st-year students. Additionally, there is a strong interference from L2 (English) to L3 (German), since the English word "central" has lexical stress on the first syllable.

Another result comes from the general comparison of pre- and post-test: Altogether, only 1 of 15 students in all three classes A, B and C had proper accentuation of this word in the pre-test. However, 11 of 15 students had proper accentuation in the post-test. Most likely, this improvement was through the impact of visualizations, notably through PW.

Even more striking is the fact that all 5 students of class B stressed the wrong syllable in the pre-test, but got it right in the post-test, though none of them used audio in either of the tests. "Love makes blind" is a common saying. Could it also be, that "Listening makes deaf"? Again, class C, which relied more on audio than on PW, improved only in 50% of the cases (2 of 4 students).

Concerning the research questions it can be said that visualizations could play an important role in teaching prosody, though in conjunction with various other techniques which have been developed for different combinations of L1, L2, and L3 (e.g. rhythmic exercises, shadowing as described in [9] for Chinese and Mongolian learners of Japanese, the verbo-tonal method by Petar Guberina as applied in [10] for English learners of French, and normal audio input). In particular in irregular cases of word stress or in cases of a strong interference of L1 or L2 onto L3, visualizations could be a valuable aid for students to stress words or utterances properly.

However, the results of this experiment cannot be generalized, since the number of students who read the same text was rather small. Therefore, more experiments of this kind should be carried out.

4. Conclusions

Visualizations seem to be a valuable tool for teaching prosody. In fact, it turned out that the phenomenon of consideration, stress placement on the German word "zentral" [tsɛnˈtraːl], was only seldom realized correctly after only listening to the textbook CD, but to a very large extent by using the visualization technique PW. This confirms findings on duration effects of PW also reported in [7].

For certain learner types (e.g. for technical students, as with the two classes A and B) as well as for certain lexical items or linguistic structures (e.g. for irregular pronunciations, or pronunciations which are subject to interference from L1 or some other language), visualizations might turn out to be advantageous. There are even indications that PW could encourage students to shorten syllables or to reduce them, which is an essential element of developing a sense for rhythm in German or English. This potential efficacy of visualizations might explain the popularity of some English textbooks using a lot of visualizations, some of them similar to PW, for teaching both, segmental and suprasegmental prosody [11].

Except for the single German word under consideration, the prosody of the third class (class C) was quite good. The longer preparation period and the availability of the textbook CD thus showed positive effects not reported here. This suggests that not a single method, but rather a suitable selection of techniques might be best for teaching prosody.

5. References

[1] Mori Y., "Differences in the phonetic realization of English rhythm between English and Japanese speakers: Focus on duration and jaw opening." in Proceedings of the 25th General Meeting of the PSJ (The Phonetic Society of Japan), 121-126, 2011. (In Japanese)

[2] Dresing, T., Pehl, T., Georgi, D. et. al., "Audiotranskription: Lösungen für digitale Aufnahme & Transkription" (solutions for digital recording & transcription). Online: https://www.audiotranskription.de, accessed on 15 Dec 2011. (Computer program)

[3] Rude, M., "Vortrag und Workshop: Funktionen der Prosodie, ein Vorleseexperiment und prosodisches Aussprachetraining", presentation at the winter conference of the Toukai section of the JGG (Japanische Gesellschaft für Germanistik), 2011. (Delivered on 3 Dez 2011)

[4] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer (Version used: 4.4.11 Current version: 5.3.03)". Online: http://www.fon.hum.uva.nl/praat/, accessed on 15 Dec 2011. (Computer program)

[5] Fujiwara, M., Katsuragi S., Motokawa Y. et. al., Start frei!, Sanshusha, 2009. (German language textbook for beginners)

[6] Ladd, D. R., Intonational Phonology, (2nd ed.) Cambridge, 75-76, 2008.

[7] Rude M., "Prosodische Schrift: Motivation, Konzept, Anwendungsbeispiele und Wirkungen.", Neue Beiträge zur Germanistik 7 (1): 140-156, 2008. (In German)

[8] Bolinger D., Intonation, Penguin, 12-14, 1972.

[9] Rongna A. and Hayashi, R., "Accuracy of Japanese pitch accent rises during and after shadowing training", in Proceedings of Speech Prosody, 6th International Conference, 2012.

[10] Alazard C., Astésano C., Billières M. et. al., "Rôle de la prosodie dans la structuration du discours: Proposition d'une méthodologie d'enseignement de l'oral vers l'écrit en Français Langue Etrangère", Actes d'IDP 09, 2009. Online: http://makino.linguist.jussieu.fr/idp09/docs/IDP_actes/Articles/alazard.pdf, accessed on 8 Feb 2012.

[11] Gilbert J. B., "Clear speech from the start: basic pronunciation and listening comprehension in North American English", Cambridge, 2001/2009.

[1] The effects of the third format, text with underlines representing primary (short underline) and secondary (long underline) stress, have not been covered in this paper.