An Online Prosodic Reading Tutor “Suzuki-kun”
—Visualization of Hidden Prosodic Information of Japanese Speech—

By taking full advantage of a prosody prediction module in a Text-To-Speech system [1] and the F0 generation model [2], we developed a prosodic reading tutor “Suzuki-kun” as one function of OJAD (Online Japanese Accent Dictionary) [3][4]. As far as we know, this is the first educational online infrastructure that automatically shows the prosodic information of any input Japanese text visually and auditorily.

For the last decade, the quality of synthetic voices has been drastically improved. Products such as talking ATMs, computer games or eBook readers with TTS engines abound in our dairy life. However, the synthetic voices and the techniques still have not been introduced effectively as educational material to support learners to acquire Japanese speech.

On the other hand, although instructors recognize the importance of prosodic trainings for many learners, they face difficulty for offering instructions. The reason is because prosodic information is not shown visually in written texts. Since native speakers automatically do the prosodic control, they are not good at being aware of the mechanism of controlling local accents and global intonations.

Therefore, we developed a prosodic reading tutor. The focal function that should be realized for the tutor is not merely reading out any input text, but visualizing prosodic information that is usually hidden as a background process when generating speech sounds from texts. Figure 1 shows the process of visualization of prosodic features of an input. Original text with punctuation marks in 1) is replaced into phonemic KANA sequences in 2). Then prosodic prediction is processed. The result is shown in 3), where a " ’ " indicates an accent nucleus, a "/" indicates an accentual phrase boundary without a pause, a "_" indicates an accentual phrase boundary with a pause inserted, functioning as an bigger intonational phrase boundary, and a "%" indicates an unvoicing operator. Without this internal prosodic prediction module, a machine cannot read aloud the original text naturally. The final presentation to users is shown in 4), where the pitch patterns are drawn smoothly utilizing the F0 model [2] and the text below the pitch patterns shows accent nucleus in red characters and unvoiced sounds in gray hatching. Synthetic voices can also be produced with variety of female/male.

Effectiveness of this tutor is shown from users’ grateful comments through the web site or its facebook, workshops done at over 60 regions so far, and more than 450,000 accesses to the site which is drastically increased after Suzuki-kun of OJAD being released after initial OJAD release on August of 2012.

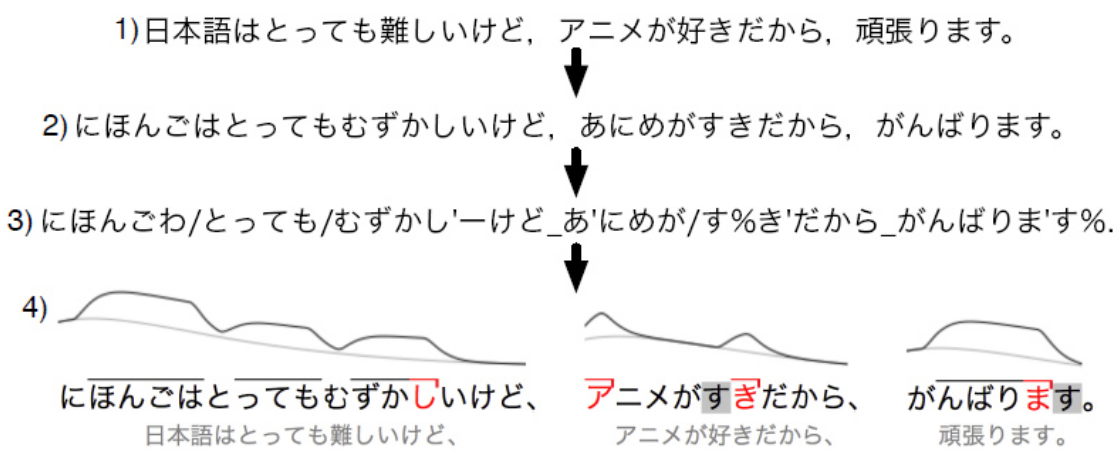

Figure1. Prediction and visualization of prosodic features for a given text.

References

[1] N. Minematsu et al, Proc. INTERSPEECH, CD-ROM, 2012
[2] H. Fujisaki et al., J. Acoust. Soc. Japan(E), 5,4, 233-242, 1984
[3] Online Accent Dictionary (OJAD) http://www.gavo.t.u-tokyo.ac.jp/ojad/
[4] I. Nakamura et al., Proc. INTERSPEECH, 2554-2558, 2013