Development of a Prosodic Reading Tutor of Japanese – Effective Use of TTS and F0 Contour Modeling Techniques for CALL –

Nobuaki MINEMATSU †, *Hiroya HASHIMOTO* † *Hiroko HIRANO* ‡, *Daisuke SAITO* †

† The University of Tokyo, Tokyo, Japan
‡ Tokyo University of Foreign Studies, Tokyo, Japan
{mine,hiroya,dsk_saito}@gavo.t.u-tokyo.ac.jp, hirano_hiroko@tufs.ac.jp

Abstract

A text typed to a speech synthesizer is generally converted into its corresponding phoneme sequence on which various kinds of prosodic symbols are attached by a prosody prediction module. By using this module effectively, we build a prosodic reading tutor of Japanese, called Suzuki-kun, and it is provided as one function of OJAD (Online Japanese Accent Dictionary) [1]. In Suzuki-kun, by using a prosody prediction module, any Japanese text is converted into its reading (Hiragana[1] sequence) on which the pitch pattern that sounds natural is visualized as a smooth curve drawn by the F0 contour generation process model [2]. Further, positions of accent nuclei and unvoiced vowels are illustrated. Suzuki-kun also reads that text out following the prosodic features that are visualized. Since releasing Suzuki-kun, the number of accesses to OJAD has been drastically increased and for the last four months, OJAD received 129,168 accesses, 58.9 % of which were from outside Japan.

Index Terms: Prosody prediction, TTS, F0 model, Prosodic reading tutor, OJAD

1. Development of a prosodic reading tutor

For the last decade, the quality and naturalness of synthetic voices has been drastically improved and it is not uncommon that those voices are presented to learners as model utterances. Generally speaking, a Text-to-Speech (TTS) engine does not read an input text directly but reads its corresponding phoneme sequence with various kinds of prosodic symbols attached by a prosody prediction module. For example, Figure 1 shows 1) an original Japanese text, 2) its phonemic transcript as Hiragana sequence, 3) output from a prosody prediction module that we developed in [3], and 4) output from Suzuki-kun. In 3), the prosodic features are predicted and represented using symbols. ' is an accent nucleus. / and _ indicate an accentual phrase boundary without a pause and that with a pause, respectively. The latter also functions as intonational phrase boundary[2]. In other words, 3) includes complete description of the hierarchical structure of prosody required to read this text naturally. 3) claims that this sentence should be divided into three intonational phrases and that, from the head of the sentence, an intonational phrase contains three, two, and one accentual phrase(es). Further, % is an unvoicing operator. Without these prosodic instructions, a machine cannot read the original text naturally.

On general textbooks of Japanese, although all the sentences have their Hiragana sequences as reading, no prosodic features are visualized and only read samples are provided as audio CD. However, it is true that only from listening, it is not easy even for native teachers to detect the hierarchical structure of prosody and the positions of accent nuclei because native speakers' prosodic control is almost unconscious and therefore, awareness of prosodic control is not always high. We can claim that only with listening to read samples, it is not rarely difficult for learners to realize natural prosody on their utterances.

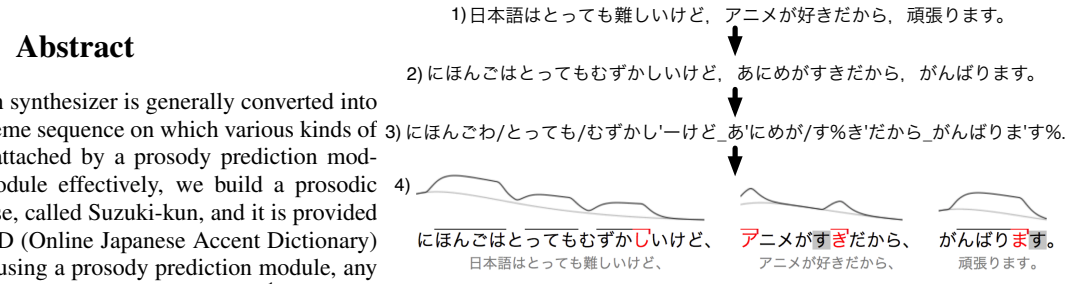

Figure 1: Prediction and easy-to-understand visualization of prosodic features for a given Japanese text

To embody the *hidden* hierarchical structure of prosody and some other prosodic features, we developed Suzuki-kun and its output is shown as 4) in Figure 1. Pitch contours are generated smoothly by the F0 model, which cover even unvoiced segments. Accent nuclei are shown in red and unvoiced morae are indicated as gray patches. Organization of intonational phrases and accentual phrases are clearly visualized. Suzuki-kun can read out texts following the visualized prosodic features.

Since releasing Suzuki-kun, the number of accesses to OJAD has been drastically increased. Now, OJAD is translated into 13 different languages. Recently, we received users' reports from China and Indonesia that almost all the finalists in Japanese speech contests practiced repeatedly using Suzuki-kun. Readers who are interested in improvement of learners' speaking performance by using Suzuki-kun should refer to [4].

2. Conclusions

By taking full advantage of a prosody prediction module in a TTS system and the F0 model, we developed a prosodic reading tutor. As far as we know, this is the first educational infrastructure of Japanese that can show the prosodic hierarchy visually and auditorily for any text. Similar infrastructure is possible for any language by using a TTS system of that language.

3. References

[1] I. Nakamura *et al.*, Proc. INTERSPEECH, 2554–2558, 2013
[2] H. Fujisaki *et al.*, J. Acoust. Soc. Japan (E), 5, 4, 233–242, 1984
[3] N. Minematsu *et al.*, Proc. INTERSPEECH, CD-ROM, 2012
[4] http://youtu.be/It-NBJKJd1g

[1] Hiragana is functionally similar to phonemic symbols of Japanese.
[2] The symbolic representation of 3) is called JEITA format in the Japanese community of Text-to-Speech synthesis.