9th ISCA Speech Synthesis Workshop
13-15 Sep 2016, Sunnyvale, USA

Prosodic Reading Tutor of Japanese, Suzuki-kun – The first and only educational tool to teach the formal Japanese –

Nobuaki MINEMATSU†, *Daisuke SAITO*†, *Nobuyuki NISHIZAWA*‡

†The University of Tokyo, Project OJAD
‡KDDI R&D Laboratories, Inc.

{mine,dsk_saito}@gavo.t.u-tokyo.ac.jp, no-nishizawa@kddilabs.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 feature of OJAD (Online Japanese Accent Dictionary) [1, 2]. In Suzuki-kun, any Japanese text is converted into its reading (Hiragana[1] sequence) on which the pitch pattern that sounds natural as Tokyo Japanese (the formal Japanese) is visualized as a smooth curve drawn by the F0 contour generation process model [3]. Further, the positions of accent nuclei and unvoiced vowels are illustrated. Suzuki-kun also reads that text out following the prosodic features that are visualized. Suzuki-kun has become the most popular feature of OJAD and so far, we gave 90 tutorial workshops of OJAD in 27 countries. After INTERSPEECH, we'll give 6 workshops in the USA this year.

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

1. Development of a prosodic reading tutor

For the last decade, the 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 [4], 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 as Tokyo Japanese (the formal Japanese). Further, % is an unvoicing operator. Without these 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,

[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.

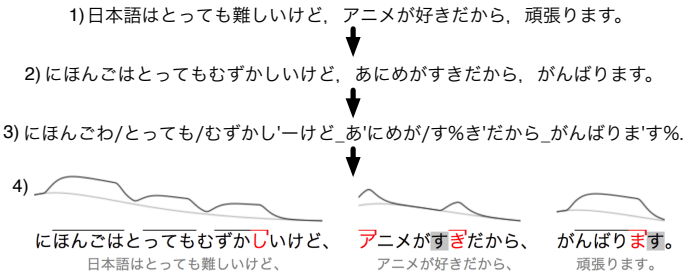

Figure 1: Prediction and easy-to-understand visualization of prosodic features for a given Japanese text

awareness of prosodic control is not always high. This is why many teachers of Japanese do NOT teach the formal Japanese to learners, whose spoken Japanese will become accented almost inevitably. With an audio CD, learners can listen to the formal Japanese, but we claim that only with listening, it is difficult for learners to realize natural prosody on their spoken Japanese.

To reveal the *hidden* hierarchical structure of prosody, we developed Suzuki-kun and its output is shown as 4) in Figure 1. Pitch contours are generated 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 prosody. Our experiments showed that visualized prosody can improve the prosodic naturalness of learners' spoken Japanese more effectively than auditory prosody, i.e. spoken samples [5].

In the long history of Japanese education, Suzuki-kun is the first and only assistive tool that teaches how to read a given text in the formal Japanese. So far, we gave 90 tutorial workshops of OJAD in 27 countries and OJAD has been translated into 14 non-Japanese languages. Beijing users claim that nobody can become a finalist of a speech contest without practicing with Suzuki-kun. Their improvements can be checked in [6].

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 of Japanese. Similar assistive tools are possible enough for any language by using a TTS system of that language.

3. References

[1] I. Nakamura *et al.*, Proc. INTERSPEECH, 2554–2558, 2013
[2] http://goo.gl/yLlqH7
[3] H. Fujisaki *et al.*, J. Acoust. Soc. Japan (E), 5, 4, 233–242, 1984
[4] N. Minematsu *et al.*, Proc. INTERSPEECH, CD-ROM, 2012
[5] N. Minematsu *et al.*, Proc. Speech Prosody, 252–256, 2016
[6] https://goo.gl/ENZog6
https://goo.gl/lKZ3co