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 intonational 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.\n\(^2\)The symbolic representation of 3) is called JETIA format in the Japanese community of Text-to-Speech synthesis.