

New Features and Effectiveness of Suzuki-kun, the First and Only Prosodic Reading Tutor of Tokyo Japanese

Nobuaki MINEMATSU, Daisuke SAITO

The University of Tokyo Project OJAD

{mine, dsk_saito}@gavo.t.u-tokyo.ac.jp

Abstract

We have updated Suzuki-kun, which is the prosodic reading tutor of Japanese that we developed as one feature of OJAD (On-line Japanese Accent Dictionary) [1, 2]. The original Suzuki-kun can visualize the prosodic and hierarchical structure of any given sentences to read them aloud in Tokyo Japanese and provide a synthetic speech sample based on the visualized prosody. The added new features are 1) generation of multiple speakers' voices, 2) speaking rate control, and 3) generation of dialogue speech among these speakers. Further, experimental results of examining the effects of Suzuki-kun on the naturalness of learners' spoken Japanese are described. It was found that visualized prosody is significantly more effective than auditory prosody. As of late August in 2017, 111 tutorial workshops of OJAD will have been given in 29 countries. In demonstration, teachers' reports will also be shown as well as Suzuki-kun's new features.

Index Terms: OJAD, Suzuki-kun, Tokyo Japanese, visualized prosody, naturalness assessment

1. Suzuki-kun, the first and only teaching material for Tokyo Japanese

Native Japanese generally speak Tokyo Japanese (TJ) in public even when they are from local cities. TJ has a unique prosodic control of word accent and phrase intonation, but it is rarely taught to learners because TJ's prosodic control is complicated and teaching time is limited. A lexical accent dictionary published for professional broadcasters [3] explains TJ to be *the dress code for speaking Japanese in public*, but it is not taught well to learners. Suzuki-kun was developed to change this problematic situation and it is based on speech synthesis technologies. For a given sentence, a speech synthesizer does text analysis at first partly for prosody prediction. In Japanese, accentual phrase boundaries and intonational phrase boundaries are predicted. Further, within a predicted accentual phrase, the position of its accent nucleus is detected automatically. Suzuki-kun uses the predicted prosody to visualize the pitch contour and the positions of accent nuclei to read the given sentence correctly in TJ. The predicted (and symbolic) prosody is also used to generate synthetic speech, which follows the visualized prosody. The schematic procedure of Suzuki-kun is shown in Figure 1.

2. New features added to Suzuki-kun

The original Suzuki-kun was demonstrated at SLaTE2015 [4], after which, we added several new features to Suzuki-kun. They are 1) generation of multiple speakers' voices (2 males and 2 females), 2) speaking rate control, and 3) generation of dialogue speech among the four speakers. For the third feature, input text has to include some command lines as shown in Figure 2. In a command line, which starts with double slashes, speaker identity and speaking rate are specified so that they are used for Suzuki-kun to read aloud subsequent sentences.

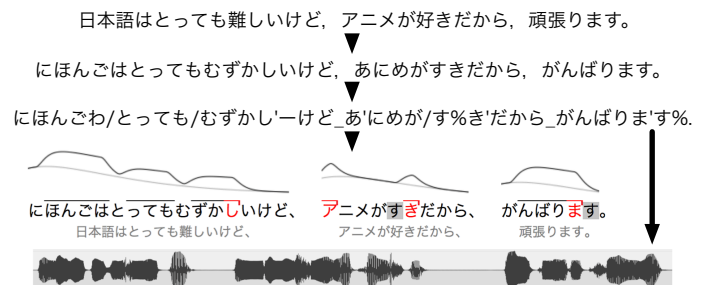


Figure 1: The schematic procedure of Suzuki-kun

```
// M1 F
来週の研究発表の準備は/できてますか？
// F2 S
パワーポイントのスライド作成に/時間がかかってます。
なので、申し訳ないですが、あと、数時間、
// F2 F
えーと、半日はかかりそうです。
```

Figure 2: A dialogue manuscript with command lines

On a command line starting with double slashes, speaker identity and speaking rate are specified explicitly.

3. Effectiveness of Suzuki-kun

Effectiveness of Suzuki-kun was verified by asking eighty Chinese learners of Japanese to practice with Suzuki-kun [5]. They were divided to three groups. Group A practiced reading a given text only with that text and then practiced reading with synthetic speech from Suzuki-kun (auditory prosody). Group B practiced only with text and then practiced with visual output from Suzuki-kun (visualized prosody). Group C practiced with both auditory and visualized prosody after practicing only with text. By comparing the prosodic naturalness observed in the first and the second practices, group C showed the largest improvement. Comparison between the two groups of A and B showed that visualized prosody was significantly more effective than auditory prosody. Detailed results are described in [5].

4. For demonstration

Suzuki-kun is the first and only pedagogical material to teach and learn the dress code for speaking Japanese in public. After releasing it, we received an enormous number of accesses especially from overseas. OJAD has been translated into 14 non-Japanese languages for beginners. As of late August in 2017, we will have given 111 tutorial workshops of OJAD in 29 countries. The latest one will be at University of Göteborg, Sweden before INTERESPEECH. In demonstration, not only Suzuki-kun's new features but also teachers' reports will be explained.

5. References

- [1] I. Nakamura *et al.*, *Proc. INTERSPEECH*, 2554–2558, 2013
- [2] <https://www.gavo.t.u-tokyo.ac.jp/ojad>
- [3] *NHK accent dictionary*, published by NHK publisher, 2016
- [4] N. Minematsu *et al.*, *Proc. SLaTE*, 189, 2015
- [5] N. Minematsu *et al.*, *Proc. Speech Prosody*, 257–261, 2016