Jiang, B. et al. (Eds.) (2025). Proceedings of the 33rd International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

Improvement of a Japanese Language Learning Support System that Enables Word Accent Learning

Kazuki TOMITA[a*], Satoru KOGURE[b], Yasuhiro NOGUCHI[b], Koichi YAMASHITA[c], Tatsuhiro KONISHI[b] & Makoto KONDO[b]
[a]*Graduate School of Integrated Science and Technology, Shizuoka University, Japan*
[b]*Faculty of Informatics, Shizuoka University, Japan*
[c]*Faculty of Business Administration, Tokoha University, Japan*
*tomita.kazuki.20@shizuoka.ac.jp

Abstract: Prosody is an important element in verbal communication. Accent, a form of prosody, plays a role in smooth communication in Japanese. In a previous study, we developed a system that visualizes Japanese word accents as accent graphs. In this study, we analyzed correct accent data and added algorithms to improve the accuracy of the accent estimation system, as well as an accent correctness judgment system. As a result, the system developed in this study was more accurate than the previous one.

Keywords: Japanese language learning, Japanese word accent recognition, accent graph

1. Introduction

In recent years, the number of foreign residents and workers in Japan has increased. To live in Japan, they must be able to communicate in Japanese. Prosody is a crucial element of spoken communication, and among its components, accent is particularly significant in Japanese because it helps distinguish word meanings. Japanese employs a pitch accent system, which plays a vital role in preventing misunderstandings and ensuring smooth communication. In previous research, we developed a Japanese dictogloss learning support system that provides an environment for acquiring Japanese pitch accents (Kogure et al., 2024). The present study aims to improve the accuracy of accent estimation and correctness judgment in that system.

2. Related Works

Matsuzaki and Takahashi (2015) addressed prosody learning systems for Japanese. In their system, learners' utterances and model utterances are visualized as “prosody graphs.” These graphs display pitch based on the prosodic features of Japanese and allow learners to compare their utterances with the model visually. However, a limitation of prosody graphs is that slight deviations in the slope between the learner's and model's graphs can mislead learners, causing them to overlook significant errors.

Kogure et al. (2024) proposed a system that visualizes pitch using an "accent graph." Like prosody graphs, accent graphs display pitch based on Japanese prosody but differ by representing pitch using only two levels: "High" and "Low." This binary representation avoids the issue of minor slope differences in prosody graphs. Therefore, the present study adopted accent graphs for pitch visualization.

The Online Japanese Accent Dictionary (OJAD) is a dictionary for Japanese learners (Minematsu et al., 2012). We used OJAD's pitch-accent information as ground-truth data, and a subset of its voice samples for verbs was employed in the evaluation experiments.

3. Japanese Pitch Accent in This Study

This study followed the Tokyo dialect accent standard. Figure 1 illustrates the accent types of four-mora words—Type 0, Type 1, Type 3, and Type 4—with each mora represented by a dot. The accent core is shown in white. In Type 4, a pitch drop occurs on the postpositional particle *ga* (the fifth mora), indicating a falling pitch after the noun. Focusing on word-level accent learning, only nouns and verbs with two or more morae are considered. To improve the accuracy of accent estimation and correctness judgment, we analyzed OJAD's reference audio data. Using *Praat*, we visualized pitch graphs for 90 audio samples and manually identified distinguishing characteristics between Type 0 and Type 1 accent words. These features were then used to enhance the system's algorithms.

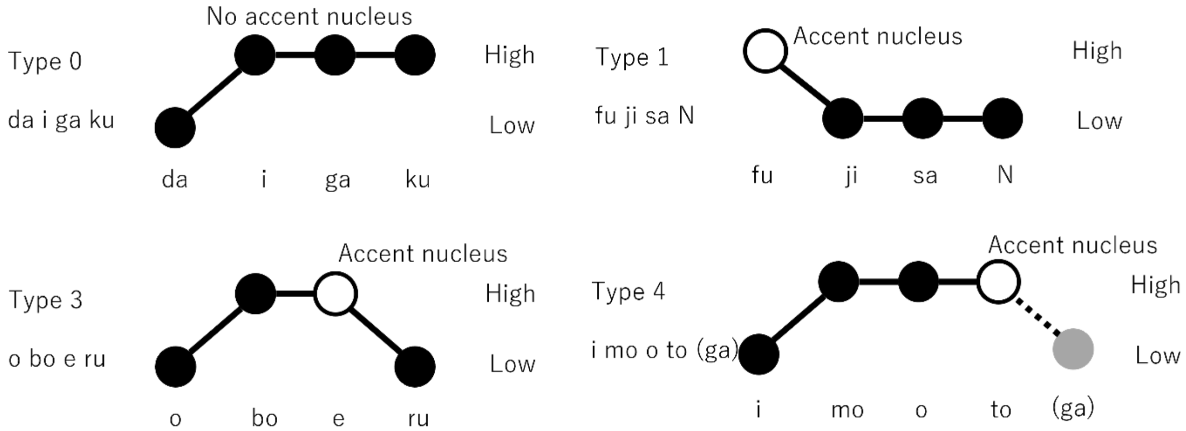

Figure 1. Examples of pitch graphs for each accent type.

4. Accent Learning System

The system estimated accent graphs based on the high and low pitch levels for each mora. The estimation algorithm was implemented using *Praat* scripts. Building on Kogure (2024), we refined the algorithm using the characteristics identified in Chapter 3 for Type 0 and Type 1 accent words. First, a reference pitch value was set. Then, the average fundamental frequency of each mora was measured and compared with the reference pitch. If the average pitch was equal to or higher than the reference, it was labeled “High”; otherwise, “Low.” The reference value was adjusted depending on the accent type: for Type 0 and Type 1, the first mora's pitch was used as the reference; for other types, the overall average pitch was used.

Correctness was assessed by comparing the estimated pitch pattern with the correct patterns derived from the OJAD and Japanese accent definitions. If the pitch values matched each mora, the results were considered correct. Additionally, based on the characteristics of Type 0 and Type 1 accent words identified in Chapter 3, we added new rules to the judgment algorithm to improve accuracy.

Feedback was provided in both text and visual formats based on the learner's audio and the system's estimation and judgment results. The feedback text indicated whether the accent was correct. The feedback image included the learner's waveform, pitch graph, mora segmentation, and both the estimated and correct accent graphs. These images were generated using *Praat* scripts. Auditory feedback was also provided: learners could listen to both their pronunciation and the correct pronunciation from OJAD, allowing for auditory comparison.

5. Evaluation Experiments

We evaluated the accuracy of the improved accent estimation and correctness judgment system. For comparison, we used the system developed by Kogure et al. (2024). The evaluation used 370 audio samples: 90 correct samples from OJAD and 280 samples (both correct and incorrect) from native Japanese speakers. Both systems performed accent estimation and correctness judgment on these samples, and we calculated the accuracy and F-measure as a comprehensive accuracy metrics.

As a result, the accuracy of the research system by Kogure et al. (2024) was 0.692 and the F-value was 0.674, whereas the accuracy of the system in this study was 0.816 and the F-value was 0.848. This improvement is attributed to the fact that the previous system often misjudged correctly pronounced accents as incorrect, which the new system successfully addressed.

6. Conclusion

In this study, by analyzing speech uttered with correct accent and enhancing the system with new algorithms, we aimed to improve the accuracy of the accent estimation and correctness judgment system developed by Kogure et al. (2024). As a result, the accuracy of accent estimation and accent correctness judgment system was improved.

We plan to conduct a usability evaluation experiment using the developed system interface. Participants will use the system and complete a questionnaire regarding their experiences. Additionally, we plan to develop a system that can support the learning of Japanese compound accents and sentence intonation by extending the learning targets.

Acknowledgements

This study was supported by JSPS KAKENHI (Grant Number: JP24K15205).

References

Boersma, P., & Weenink, D. (2024). *Praat: doing p honetics by computer* (Version 6.2.06) [Software]. https://www.fon.hum.uva.nl/praat/

Immigration Services Agency. (2024). *Number of foreign residents as of the end of June 2024*. https://www.moj.go.jp/isa/publications/press/13_00047.html (in Japanese)

Kogure, S., Tomita, K., Noguchi, Y., Yamashita, K., Konishi, T., & Kondo, M. (2024). Construction of a Japanese language learning support system that enables word accent learning. *Proceedings of the 31st International Conference on Computers in Education* (ICCE 2023) (pp. 329-334)

Lee, A., & Kawahara, T. (2019). *Julius* (Version 4.5) [Software]. https://doi.org/10.5281/zenodo.2530395

Matsuzaki, H., & Takahashi, E. (2015). Development of a system for automatic output of prosody graphs corresponding to a delayed fundamental frequency fall. *Journal of Japanese Language Education Methods, 22*(1), 48-49 (in Japanese). https://doi.org/10.19022/jlem.22.1_48

Minematsu, N., Kobayashi, S., Shimizu, S., & Hirose, K. (2012). Improved prediction of Japanese word accent sandhi using CRF. *Proceedings of Interspeech 2012 ISCA's 13th Annual Conference* (pp. 2562–2565). https://doi.org/10.21437/interspeech.2012-663

Ministry of Health, Labor, and Welfare. (2024). *Summary of reporting status of "Employment of foreign nationals" (as of the end of October 2023)*. https://www.mhlw.go.jp/stf/newpage_37084.html (in Japanese)

The Japan Foundation. (2009). Japanese language teaching methods series, Vol. 2: Teaching Voice. The Japanese Foundation. (in Japanese)