Automated judgment of Japanese pitch accent from speech

Background and overview

The Japanese language uses pitch accent as a distinctive feature in its morphology: two words that are otherwise pronounced the same but differ in pitch contour can have distinct meanings. Japanese Pronunciation Pro is a website project led by Nishimura-sensei of the Japanese department which aims to provide an introduction of this topic for learners of Japanese. Currently the website consists primarily of instructional materials including explanatory text, audio examples and pitch diagrams generated automatically from the examples.

One of the features we want to add to this site is an automated speech practice/quiz system that can record a student's attempt to pronounce a word with the correct pitch accent and determine if it was pronounced correctly. Using the existing set of examples, which have the expected pitch contours associated with them, we can then automatically create quizzes and practice sessions for the instructional pages based on the sets of examples used in them–this kind of active learning could prove much more helpful for students than simply reading the material on the site.

Prior research and resources

There exists a great deal of research on the specific problem of pitch accent in Japanese. Kawai and Ishi (1999) [1] develop a model to categorize pitch accent contours based on a fundamental frequency which they use for speech synthesis. Hirose (2004) [2] applies the aforementioned work to develop a system for detecting pitch accent in speech for the specific case of a pitch accent training program.

However, with more modern NLP tools, a simpler approach might also be worth exploring. With tools like allosaurus that can provide phoneme extraction, we could build a system that slices up an audio sample based on the detected phonemes, and measures the average pitch of each of these slices, using a model similar to Kawai and Ishi's to determine a pitch contour which can then be compared to our expected shape. Libraries such as CMUSphinx provide a host of other speech recognition and processing tools that might prove useful.

Should we wish to train a machine learning classifier for some or all portions of this pipeline, we can generate a host of examples relatively quickly because Nishimura-sensei and some of her past students have recorded hundreds of examples of pronunciations of Japanese words, and we can label them with the correct contour easily since they are mostly categorized already.

Project plan

The first phase of the project is to perform a deeper investigation of the existing literature and techniques previously used for this or related tasks. I then plan to implement a prototype of the system and test it on the examples we already have as well as my own speech. I hope to iterate and refine this system outside of the context of the website for a few weeks before moving on to the second phase, which will be implementing a web interface for it and connecting it to the website. I plan to work with the Japanese department and perform some student tests to design a good interface for this, after which we can automatically generate content for quizzes and practice sessions based on the relevant words for each instructional section of the site.

Timeline

Date	Description	Details
Feb 10	Complete literature discovery and summary	Find relevant papers for speech recognition and/or synthesis and summarize important findings, including techniques for audio processing, models of pitch contour, and normalizing between different speakers
Feb 24	Initial prototype of detection system	Implement in a CLI application; show evidence of detecting and segmenting speech into morae and obtaining some pitch contour
Mar 31	Final, refined version of detection system	This version can reliably detect the pitch accent of the existing speech examples and verify them as correct, as well as work on novel recordings of my own speech and that of a native speaker
Apr 14	UI design for website integration	This design will be developed iteratively with feedback of Nishimura-sensei and her students and with inspiration from other language pedagogy websites
Apr 28	Integration with live instructional pages; automatic content generation	The quiz and practice modes of JPP are usable and can automatically generate content from the example words in a given section

References

[1] Kawai, Goh, and Carlos Toshinori Ishi. "A System for Learning the Pronunciation of Japanese Pitch Accent." 6th European Conference on Speech Communication and Technology (Eurospeech 1999), ISCA, 1999, pp. 177–82. DOI.org (Crossref), https://doi.org/10.21437/Eurospeech.1999-48.

[2] Hirose, Keikichi. "Accent Type Recognition of Japanese Using Perceived Mora Pitch Values and Its Use for Pronunciation Training System." (2004). https://www.isca-speech.org/archive_open/tal2004/tal4_077.pdf.