

Development of an L2 Japanese speech corpus for the comparison of prosody across diverse L1 groups

Motoko Ueyama¹ Ryoko Hahashi²
Aaron Lee Albini²

¹University of Bologna, ²Kobe University

Prosody is a significant contributing factor in how native speakers rate the pronunciation of second language (L2) learners [1]. Compared to individual segments, learners' divergences from target-like prosody have been reported to have a stronger effect on native speakers' evaluations [2] and lead to more serious communication difficulties [3]. In the case of L2 Japanese, prosody is known to be one of the most challenging aspects of Japanese pronunciation for learners to acquire [4]. To better understand which specific aspects of prosody pose the greatest problems (and therefore are of the greatest pedagogical importance), there is a need for large-scale databases of experimentally elicited samples of learner speech [5]. Unfortunately, such resources are still scarce. While a handful of such corpora do exist for L2 Japanese (see [6]), none of them explicitly target a wide range of prosodic phenomena while simultaneously sampling from a diverse set of native-language (L1) groups.

The present study reports on an on-going corpus construction project to fill this gap. At present, the corpus consists of Japanese learners mainly at the upper elementary level, representing four different L1 backgrounds: Chinese (N=29), Korean (N=23), Italian (N=16), and Russian (N=15). In addition, a sample of 9 native speakers of Tokyo Japanese is also included as a baseline. The stimulus materials were designed to examine five aspects of Japanese sentence-level prosody: (1) final rise intonation in interrogatives, (2) focal prominence, (3) downtrends (especially downstep), (4) the disambiguation of syntactically ambiguous sentences, and (5) the prosodic realization of emotions and other paralinguistic information. To serve as a point of reference for comparing the acquisition of prosody vs. segmental phenomena, a sixth set of materials was also included testing vowel devoicing. Data is being collected using the Online Voice Recorder tool referenced in [7]. Upon completion of the project, the corpus will be made publicly available online.

A preliminary analysis comparing the L1 Italian data to the native Tokyo Japanese baseline revealed learners to diverge from native norms in the following ways:

1. **Interrogatives:** Learners often inserted a superfluous pitch accent on the penultimate syllable of lexically unaccented words, but only when produced in interrogatives (i.e., not in the parallel declaratives).
2. **Prominence:** Many learners (A) produced sentences with no prominence at all, (B) placed prominence on an unfocused word, or (C) realized prominence by inserting a superfluous pitch accent on a lexically unaccented word.
3. **Downtrends:** Several learners (especially less fluent ones) failed to produce the native-like stairstepping sequence of downsteps.
4. **Disambiguation:** Many learners' strategies involved inserting a long pause and/or dividing the sentence into multiple IPs (in contrast to the native speakers, who generally phrased each sentence as a single IP).
5. **Paralinguistics:** While learners could produce *joy* and *surprise* in native-like ways, this was not true for *anger*, *sadness*, and *seriousness*.
6. **Devoicing:** None of the learners had difficulty devoicing the relevant vowels.

The unique design of the corpus makes two kinds of comparison possible. First, through analyses like those just described, each individual L1 group can be examined in terms of which of the six phenomena are most challenging. Second, the various L1 groups can be compared in terms of their performance with respect to one specific phenomenon (e.g., downstep). The results of such analyses hold great potential not only for L2 prosody research but also for the future development of tools and materials for teaching L2 Japanese.

References

- [1] Lengeris, A. (2012). Prosody and second language teaching: Lessons from L2 speech perception and production research. In J. Romero-Trillo (Ed.), *Pragmatics and Prosody in English Language Teaching* (pp. 25-40). Dordrecht, Netherlands: Springer.
- [2] Sato, T. (1995). Tan'on to inristu ga nihongo onsei no hyōka ni ataeru eikyō. *Sekai no Naka no Nihongo Kyōiku*, 5, 139-154.
- [3] Celce-Murica, M., Brinton, D. & Goodwin, J. (1996). *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. Cambridge, UK: Cambridge University Press.
- [4] Toda, T. (2008). Hatsuon no tatsujin to wa dono yō na gakushūsha ka. In Toda, T., (Ed.), *Nihongo Kyōiku to Onsei* (pp. 61-80). Tokyo, Japan: Kuroshio Shuppan.
- [5] Gut, U. (2014). Corpus phonology and second language acquisition. In J. Durand, Gut, U., & Kristoffersen, G. (Eds.), *The Oxford Handbook of Corpus Phonology* (pp. 286-301). Oxford, UK: Oxford University Press.
- [6] Maekawa, K. (2009). Nihongo gakushūsha onsei kenkyū no kadai. *Nihongo Kyōiku*, 142, 4-13.
- [7] Yue, J., Shiozawa, F., Toyama, S., Yamauchi, Y., Ito, K., Saito, D. & Minematsu, N. (2017). Automatic scoring of shadowing speech based on DNN posteriors and their DTW. In *Proceedings of Interspeech 2017* (pp. 1422-1426). DOI: 10.21437/Interspeech.2017-728.