# Perceptual Cues of Japanese /r/ Sounds: Formant Transitions vs. Intensity Dip

*Takayuki Arai*

Sophia University
arai@sophia.ac.jp

Japanese /r/ has several allophones (Kawakami, 1977; Hattori, 1984; Vance, 1987; Saito, 1997), and is typically pronounced as a flap, especially in the intervocalic position. While phrase-initial /r/ is often produced as a plosive-like sound, approximants (including lateral approximants) are another type of allophonic variation that is also frequently observed in real speech. A study of spontaneous Japanese pointed out that the variations were more widely distributed than speech observed in laboratory settings (Arai, 1999). A study of children's speech also reported that the acquisition of /r/ in Japanese is delayed compared with other consonants (Arai, 2013). The same study looked at a wide variety of phones as allophones for real Japanese /r/ sounds: ɖ (retroflex plosive), ɾ (alveolar flap), ɽ (retroflex flap), ɹ (alveolar approximant), l (alveolar lateral approximant), ɭ (retroflex lateral approximant), and ɺ (alveolar lateral flap). The study involved a perceptual experiment, where the likeliness of Japanese /r/ was asked to listeners for each /ara/ stimulus synthesized based on the perceptual cues of formant transitions and a short dip in intensity covering widely distributed sounds in real speech, but only limited discussions were developed.

In the present study, we, therefore, further categorized the experimental results in Arai (2013) related to the allophonic variations in real Japanese /r/. Although the third formant movement plays an important role, especially for the retroflex sounds, Arai (2013) mainly focused on the first (F1) and second (F2) formant transitions and the intensity dip for targeting alveolar sounds. Figure 1(a) shows a schematic plot of the four parameters used to synthesize variations in /ara/ stimuli: depth of the intensity dip, or AVd; duration of the intensity dip, or AVg; total duration of the formant transitions, or Fg; and peak of the F2 frequency, or F2p. The AVd was varied from 0 to -40 dB in 10-dB decreasing steps. The AVg and Fg were independently varied from 20 to 100 ms in 20-ms steps under the condition of Fg >= Ag. The F2p was varied from 1200 to 1600 Hz in 200-Hz steps. A set of 195 /ara/ stimuli in total generated using a formant synthesizer was presented to twenty listeners, who judged the likeliness of Japanese /r/ on a 0-100% scale.

**Figure 13: (a) Schematic plot of four parameters as function of time** used to synthesize 195 /ara/ stimuli in Arai (2013). (b)-(d) Magnified plots of parameters of stimuli judged as /r/-like sounds with high scores.
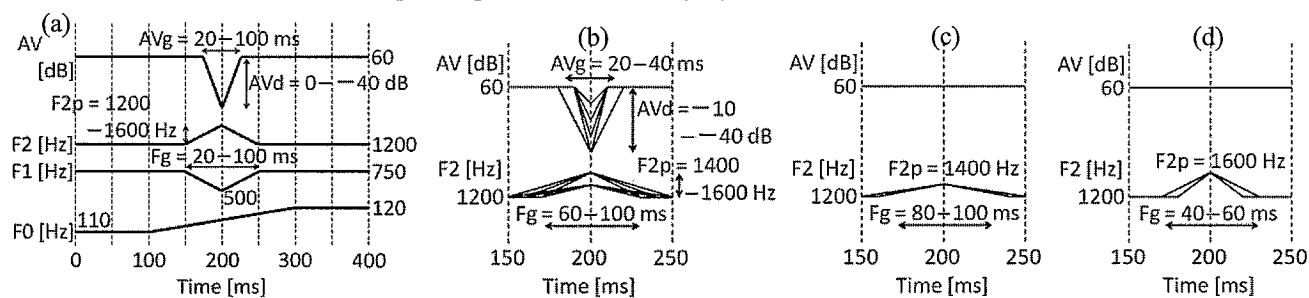


In the present study, we categorized the results having high scores of 70% or greater in terms of the likeliness of /r/. The patterns are summarized as diagrams in Figs. 1(b)-(d). In Fig. 1(b), the intensity dip along with the formant transitions yielded a clear alveolar flap sound, which is a typical phone in the intervocalic context. When there is no F2 transition, that is, F2p = 1200 Hz, the resultant stimuli were only judged as Japanese /r/ with low scores. This suggests that the F2 transition towards 1400-1600 Hz acts as crucial cue reflecting the tongue movement against the alveolar ridge. In addition, the intensity dip induces a sensation of the flap sound. In Fig. 1(c), there was no short dip in intensity, but the stimuli sounded an alveolar lateral approximant, and most listeners judged them as a natural Japanese /r/. When the duration of the F2 transition was short enough, such as approximately 50 ms, the resulting stimuli sounded an alveolar lateral flap, which was the case in Fig. 1(d), and also sounded a clear Japanese /r/. In conclusion, the results suggest that continua for real Japanese /r/ have multiple regions where the sounds can be identified as Japanese /r/ and such regions correspond to "sub-phonemes" of real /r/. In other words, Japanese native listeners might perceive such sub-phonemes categorically, and peaks can be observed around the sub-phonemic boundaries in a discrimination test.

# References

Arai, T. 2013. *Proc. INTERSPEECH Lyon.* / Arai, T. 1999. A case study of spontaneous speech ·in Japanese. *Proc. ICPhS San Francisco*, 615-618. / Hattori, S. 1984. *Onsei-gaku.* Iwanami. / Kawakami, S. 1977. *Nihongo Onsei Gaisetsu*, Ohfu. / Saito, Y. 1997. *Nihongo Onsei-gaku Nyumon.* Sanseido. / Vance, T. J. 1987. *An Introduction to Japanese Phonology.* State University of New York Press.