# Formant-altered auditory feedback on non-native vowel production

Sadao Hiroya and Takemi Mochida
NTT Communication Science Laboratories, NTT Corporation, Japan

Auditory feedback while speaking plays an important role in stably controlling speech articulation. Its importance has been verified in formant-altered auditory feedback (AAF) experiments in which speakers utter while listening to speech with a perturbed formant frequency [1]. However, in many formant AAF experiments, auditory feedback in speaking the native language has been investigated, and few experiments have been performed in speaking the non-native languages.

In this study, we conducted a formant AAF experiment for native Japanese speakers, which converts the vowels of English syllable “had” or Japanese mora (syllable-like phonological unit) “ha” to English vowel sounds [æ]. Note that native Japanese speakers tend to transfer the English vowel [æ] to the Japanese vowel “a” because the English vowel [æ] does not exist in Japanese. Therefore, the vowel productions of “ha” and “had” in native Japanese speakers are almost the same although they learn English as a foreign language.

Nine native Japanese speakers (seven females) participated in the experiment. All subjects had no experience of studying or living abroad. There were blocks for uttering “had” or “ha”. The blocks were in random order. The subjects wore headphones and uttered the letters as soon as they appeared on the display. One block consisted of 140 trials with no perturbation during 1-20 trials (Baseline), with the perturbation amount linearly shifted towards the formant frequency of the [æ] during 21-70 trials (Ramp), with the maximum perturbation amount maintained during 71-90 trials (Hold), and with no perturbation during 91-140 trials (Release).

The clamp condition (i.e., constant values) [2] was used for the formant frequency transformation instead of the shift condition (e.g., +100 Hz) which is usually used in AAF experiments. The transformed formant frequency at Hold was the same [æ] for both “had” and “ha”, i.e., speaking "had" and "ha" sound like h[æ]d and h[æ], respectively. The difference between these tasks was the target on the display. “ha” was displayed in Japanese characters. Based on the average values of Japanese vowels of native Japanese speakers and English vowels of native English speakers, F1 (the first formant frequency) of [æ] was taken as 100 Hz plus the value of F1 of the Japanese vowel “e” of the speaker, and F2 of [æ] was taken as the value of F2 of the Japanese vowel “e”. We used the phase equalization-based autoregressive exogenous model (PEAR) in our AAF system for its high formant estimation accuracy [3,4].

Fig. 1 shows the change in formant frequency of Hold relative to the Baseline. In “had”, the formant frequencies significantly changed in both F1 and F2. On the other hand, no

significant changes were observed in F1 and F2 in “ha”. Fig. 2 shows the change in formant frequency in the “had” at Release. In Release, no perturbation was given, but in F2, the formant frequency change was even larger in the perturbed direction in 101-130 trials than in Hold. This means that the pronunciation of [æ] has been learned. The results for “had” cannot be explained by the compensatory response to the perturbation.

Since “ha” is the native language for native Japanese speakers, a model of speech production has been established in the brain and stable speech production can be made by feedforward system. The fact that there is no change for the native language supports the result in the clamp condition of [2]. On the other hand, since “had” is the non-native language, a model of speech production has not been well established, and it is somewhat likely that speech production was affected by auditory feedback.

In conclusion, our results suggested that speaking a non-native language may be more affected by auditory feedback than native language. This finding may be useful for learning pronunciation of foreign languages.

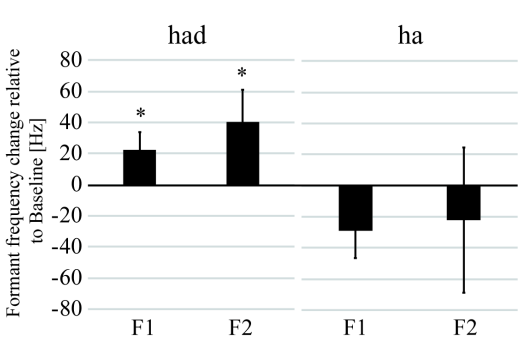


Fig 1. Formant frequency change at Hold relative to Baseline. *$p < 0.05$. The error bar is S.E.

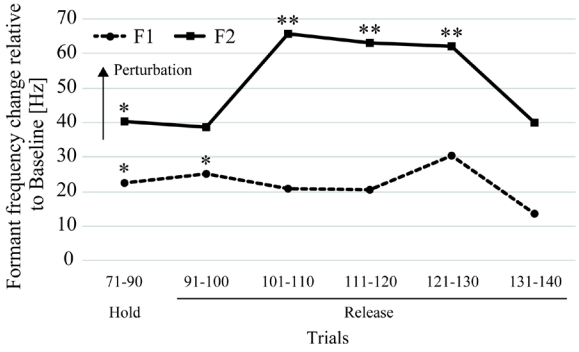


Fig 2. Formant frequency change at Hold and Release relative to Baseline for “had”. **$p < 0.01$.

## References

[1] Houde, J. F., and Jordan, M. I. (1998). “Sensorimotor adaptation in speech production,” Science 279(5354), 1213-1216.

[2] Daliri, A., and Dittman, J. (2019). “Successful auditory motor adaptation requires task-relevant auditory errors,” Journal of Neurophysiology 122(2), 552-562.

[3] Oohashi, H., Hiroya, S., and Mochida, T. (2015). “Real-time robust formant estimation system using a phase equalization-based autoregressive exogenous model,” Acoustical Science and Technology 36(6), 478-488.

[4] Uezu, Y., Hiroya, S., and Mochida, T. (2020). “Vocal-tract spectrum estimation method affects the articulatory compensation in formant transformed auditory feedback,” Acoustical Science and Technology 41(5), 720-728.