Can Japanese speakers compensate for coarticulation due to /l/ and /r/?

INTRODUCTION: Context plays a crucial role in speech perception. For example, Mann (1980) shows that when presented with a continuum changing from [d]-[g], English listeners identify more of the continuum as [g] when preceded by [al] than [ar]. Mann (1980) interpreted this effect as follows: when the target stop is preceded by [al], listeners assume that it is fronted due to coarticulation with [l], compensate for that coarticulation, and judge that consonant to be pronounced as more back, i.e., [g]. This effect is known as *compensation for coarticulation*. A follow-up experiment (Mann 1986) reported a rather surprising result: Japanese speakers, who generally cannot distinguish [r] and [l] (Goto 1971), show the compensation for coarticulation due to [r] and [l], just like English speakers. Why would Japanese speakers compensate for coarticulation, when they cannot distinguish [r] and [l], i.e. when they cannot tell what consonant precedes the target stop? In general, if Mann's (1986) results and interpretations are right, then **compensation for coarticulation should be independent of listeners' phonological knowledge** (cf. Lotto & Kluender 1998; Kingston et al. 2014 who argue that this perceptual context effect comes from a general auditory mechanism, which is language-independent). However, given recent body of evidence to the contrary (e.g. Yu et al. 2013), this conclusion should be reexamined.

Against this general theoretical background, this project re-examines Mann's (1986) finding in several respects. First, Mann (1986) did not examine individual differences, and hence a question remains as to whether all Japanese speakers can compensate for coarticulation due to [r] and [l]. Second, relatedly, Mann (1986) did not measure each listener's ability to perceive [r] and [l], and therefore it remains unclear whether the effect of compensation for coarticulation correlates with their ability to perceive the [r]-[l] distinction. Finally, Mann (1986) used naturally produced stimuli for [al] and [ar] and synthesized stimuli for the [d-g] continuum, which could have caused some unnaturalness in the overall stimulus sounds.

METHOD: The stimuli were of the following form [aXYa], where X is a {r-l} continuum and Y is a {d-g} continuum (Kingston et al. 2014). The entire stimuli were created using the Sensimetrics implementation of the KLSYN88 terminal analogue synthesizer. The two vowels are always identical, [a] with F3 of 2500 Hz. A liquid continuum {r-l} was created by incrementally varying F3: for the [ar]-endpoint, it fell to 2000 Hz, and for the [al]-endpoints, it rose to 2800 Hz. The continuum was created with 6 step increments. Poles and zeros were also manipulated following Stevens's (1998) descriptions of [r] and [l]. The liquid portion was followed by a 95 ms gap with low-frequency periodic energy to mimic closure voicing of [d] and [g]. The [d]-[g] continuum was created by varying F3: in the [da] endpoint, F3 began at 2690 Hz, while in the [ga] endpoint, it began at 2104 Hz, again with 6 step increments. There are therefore 49 stimuli (7-step {r-l} continuum times 7-step {d-g} continuum.)

In the listening phase, listeners heard one stimulus and were asked to judge whether the second syllable was [da] or [ga]. The order of the stimuli was randomized within each block. All listeners went through 8 blocks. This phrase was followed by a self-time break. In the second phase of the experiment, the listeners were presented with the [ar] and [al] endpoint stimuli in isolation, and were asked to identify these sounds. D-prime was calculated for each listener as a measure of their ability to perceive the difference between [r] and [l]. 30 native speakers of Japanese participated in this study.

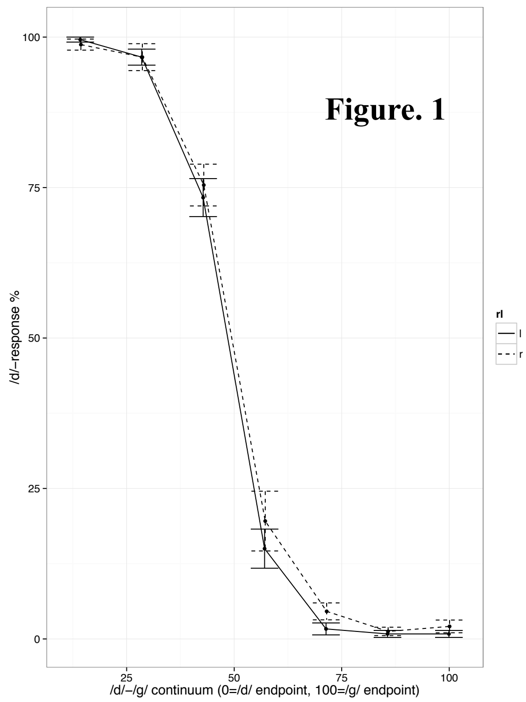

Figure. 1

RESULTS: Figure 1 illustrates the averaged identification function for the [d]-[g] continuum in the [r] and [l] endpoint environments. As with Mann (1986), in general, Japanese speakers identified more of the same {d-g} continuum as [d] after [ar] than after [al], but the effect size seems small.

Various logistic regression models were fit to the obtained responses, and the best fitting model was chosen based on Akaike Information Criterion (AIC). The best model includes the effect of the {d-g} continuum and the {r-l} continuum and its interaction as fixed factors and subjects as a random factor. The effect of the {r-l} continuum on the [d]-[g] judgment, however, was not significant in this model (z=-1.69, p=0.09).

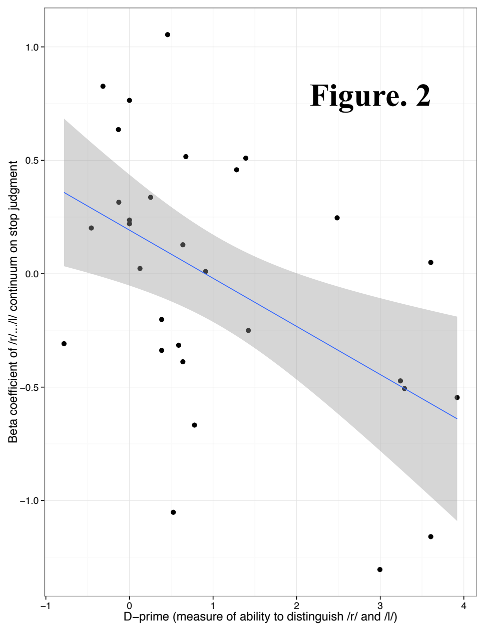

Figure. 2

To further examine the behavior of individual listeners, Figure 2 shows a scatterplot showing the correlation between the magnitude of the context effect and each listener's ability to identify [r] and [l]. The magnitude of the context effect is quantified as the beta coefficient of the [r-l] effect in the logistic regression model described above. The positive values represent that listeners identified the stop more likely as [d] after [r], as expected from Mann (1986), and the negative values show that the listeners show opposite effect. There is a significant negative correlation between the two (r=-0.50, p<.001): the better listeners were able to distinguish [r] and [l], the more likely they judge the continuum as [d] after [l] (not after [r]). This result shows that those speakers who can better perceptually distinguish [al] and [ar] are more likely to assimilate with—rather than compensate for—the context when identifying a stop preceded by a liquid.

DISCUSSION: Our study reveals several new findings going beyond Mann's results. First, not all Japanese listeners showed the compensation for coarticulation effect, as expected from Mann's (1986) study. There were 3 listeners who were insensitive to the context effect (those whose x-values are 0). Moreover, there are a fair number of listeners who showed high d-prime values—who could distinguish [r] and [l] well—and showed an "anti compensation for coarticulation effect". In general, it seems hard to conclude from our results that compensation for coarticulation is a universal mechanism, as envisioned by Mann (1986). Future theories of speech perception need to account for three questions from the current experiment; (i) where do individual differences come from?; (ii) how does an "anti compensation for coarticulation effect" arise? (ii) where does the negative correlation in Figure 2 comes from?