Perception, mental representations and production of non-native prosodic contrasts

Yuki Asano
University of Tübingen, Germany
yuki.asano@uni-tuebingen.de

Despite documentation of second language (= L2) learners’ deviant prosodic productions (e.g. Trouvain and Gut, 2007), little is still known about where such deviant productions in L2 prosody come from, in other words, which stages in L2 processing contribute to the difficulties. The current study aims to localise possible sources of difficulties and investigates the relationships between L2 perception, mental representations and production when processing non-native prosodic contrasts, in particular consonant length contrasts. A perception experiment and a production experiment were conducted with the same participants using the same stimuli.

First, 48 German learners of Japanese, 24 German non-learners and 24 Japanese first language (= L1) speakers participated in a speeded AX task with short and long inter-stimulus-intervals (= ISIs, 300 ms. vs. 2500 ms) to discriminate consonant length contrasts (in comparison to vowel length contrasts). Only consonant length contrasts were non-native for German participants. Stimuli were six disyllabic nonsense word triplets (punu, gunu, gupu, gubu, zusu, sufu, each varied into singleton, geminate and long-vowel). For same trials, two tokens of the same type were used. Following the theory of working memory (Baddeley and Hitch, 1974; Baddeley et al., 1998), the task with short ISIs tested speech perception without necessarily requiring mental representations, because the acoustic correlates of stimuli were still available. The task with long ISIs, on the contrary, required participants to access mental representations after the decay of acoustic information. Therefore, a performance decrease in terms of lower sensitivity and longer reaction times by non-native speakers was predicted in the long ISI condition due to the lack of consonant length contrasts in German. Additionally, the stimulus pairs were presented either in a flat or a falling pitch that occurred simultaneous to the segmental length contrasts. The task-irrelevant falling pitch movement increased the demand on attention control and tested how vulnerable L2 perception processing can be to distracting factors. Participants’ sensitivity to the contrasts was calculated using d’ scores (Macmillan and Creelman, 2005).

The analysis with linear mixed effects regression models (LMER) showed a significant three-way interaction between language group, ISI and pitch (p < 0.001). Splitting the data by pitch, there was an interaction between language group and ISI in the flat pitch condition (d’ scores of non-learners decreased in the long ISI condition in comparison to those of Japanese native speakers, β= -0.70, SE = 0.20, t = -3.5, p <0.001, and to those of learners, β= -0.72, SE = 0.17, t = -4.1, p < 0.001). In the falling pitch condition, there was a main effect of language group (d’ scores of non-learners were lower than those of Japanese, β= -0.93, SE = 0.20, t = -4.6, p < 0.001, and of learners, β= -0.63, SE = 0.17, t = -3.6, p < 0.001, and d’ scores of learners tended to be lower than those of Japanese, β= -0.30, SE = -1.7, p = 0.09). A very high discrimination ability was shown in all groups when task demands were lowest (= short ISI and flat pitch). With increasing task demands, only the non-natives’ discrimination abilities decreased: non-learners were strongly affected by both ISI and pitch, while learners only by pitch. Taken together, non-learners without exposure to an L2 encountered difficulties once mental representations were required. Moreover, the results showed that L2 learners were establishing novel phonological representations, but the ability to use them was demonstrated only under favourable listening conditions with no distracting acoustic information (i.e. without a task-irrelevant pitch).

After the perception experiment, the same participants took part in an immediate vs. delayed imitation task. The two imitation conditions were based on the same theoretical backgrounds as the two ISI conditions in the perception experiment. Following the theory of working memory, only the non-native speakers’ imitation accuracy of consonant length contrasts was expected to decrease in the delayed imitation condition (delay duration = 2500 ms), because the immediate imitation condition would not necessarily involve speakers’ mental representations, but the delayed imitation condition would. According to the theory of direct realism (Fowler, 1986),
speech production is driven by reflexive phonetic gestures that are mediated automatically in speech perception without requiring access to speakers’ phonological representations. If this claim is valid, the lack of non-native phonological representations should not impede sound imitation with non-native phonological structure, and non-native speakers’ imitation accuracy should not differ from that of L1 speakers in both the immediate and delayed imitation conditions. To ensure that non-native speakers could rely solely on the acoustic echo of the stimuli in the immediate imitation condition without having any difficulties in articulation, only the stimuli that were articulatorily not difficult for both Japanese and German participants were selected. Moreover, the investigation of segmental length contrasts excluded difficulties related to phonetic settings (e.g., Honikman, 1964; Mennen et al., 2010), in other words, difficulties regarding how to configure the vocal apparatus (such as the lips, tongue and jaw) for language-specific habits. The imitation accuracy was defined as how similar productions were to the stimulus words. More precisely, duration ratios of the critical consonants of the stimuli were subtracted from those of the participants’ productions.

The LMER-analysis on consonant ratios showed a significant main effect of language group (ratios produced by learners were smaller than those produced by Japanese, β= -0.42, SE = 0.05, t = -8.7, p < 0.001, ratios produced by non-learners were smaller than those produced by Japanese, β= -0.58, SE = 0.06, t = -10.6, p < 0.001, ratios produced by non-learners were smaller than those produced by learners, β= -0.17, SE = 0.05, t = -3.5, p < 0.001). Both learners and non-learners produced smaller consonant duration ratios than L1 speakers already in the immediate imitation task, suggesting that even the immediate imitation condition may have involved the speakers’ mental representations. The claim made by the direct realist view was not supported in the data.

Using the simplest discrimination and imitation task with low task demands, the study traced back to the basis of speech perception and production. The results are summarised as follows: First, both learners and non-learners were able to perceive non-native prosodic contrasts as well as L1 speakers when task demands were lowest, simply relying on the acoustic correlates of stimuli. Second, once phonological representations were required, the perception sensitivity by non-learners decreased due to the lack of phonological contrasts in their L1. Third, even the learners who were establishing L2 phonological representations showed difficulties in automatizing their L2 perception processing, easily affected by distracting factors. This result is supported by the Automatic Selective Perception model (Strange, 2009). Non-native listeners’ reduced sensitivity under increasing task demands appears to explain why even advanced learners still face difficulties in natural learning situations. Finally, both learners and non-learners showed deviant L2 productions already in the imitation condition with lowest task demands, suggesting that speech production inevitably requires access to a speaker’s phonological representations. Returning to the initial research question, the study supports the claim that mental representations cause non-native speakers’ difficulties when processing L2 prosody.

References