The Effects of Talker Variability on the Perception of American English /r/ and /l/ by Japanese Subjects: Normalization or Criteria-setting? *

James S. Magnuson and Reiko A. Yamada (ATR Human Information Processing Laboratories)

概要

　日本人が米語の /r//l/ 音を同定する際，話者によってはその話者一人の声のみが呈示されるブロックに比べ複数話者の声が呈示されるブロックにおいてより /r/ らしく，またはより /l/ らしく聞こえる場合があることを以前報告した [3]。本稿では，刺激セットによって反応パターンが変化する現象について自然音声を刺激音とし，ブロック中に出現する話者 (単一か複数か) およびブロック内の /r/ 刺激と /l/ 刺激の出現頻度を変動させることにより，話者への適応および /r//l/ の判断基準設定という問題について検討を行った。その結果，"R"と答える率は /r//l/ の出現頻度に影響されなかった。また，"R"の反応率はブロック内で呈示される話者が単一か複数かという条件の差により有意に変動した。これらの結果は"R"か"L"かという判断において，日本人は声の特徴の変化に適応するというより，むしろ "R-L"の音響的手掛かりの範囲を"R"と"L"が同数程度含まれるように２分割するよう判断基準を設定しなおすという以前の結果 [6, 3] を再確認した。

1 Introduction

Yamada and Tohkura [6] found that when Japanese subjects are forced to choose between "R" and "L" responses, they seem to set criteria for their decisions based on the *range* of stimuli they hear in a block of trials. When Yamada and Tohkura presented subjects with only the /r/-side of a synthesized /r/-/l/ continuum, subjects classified many more of the /r/-side stimuli as "L" than when they heard the /r/-side stimuli presented with the full stimulus set. Yamada and Tohkura concluded that Japanese subjects perceive /r/ and /l/ continuously, and when they are forced to choose between "R" and "L" responses, they attempt to find criteria for their decisions that result in approximately balanced numbers of "R" and "L" responses.

Previously [3], we have reported experiments in which we measured Japanese subjects' rate of "R" response (how often they respond "R" on all trials, including incorrect responses) when they listened to stimuli produced by a single talker, and when they listened to stimuli produced by several talkers mixed in random order. While the overall rate of "R" response did not change between talker conditions, the rate for some talkers increased or decreased significantly between talker conditions: some talkers were "R-like," in that subjects were more likely to label their productions of both /r/ and /l/ as "R" when their productions were mixed with other talkers' productions, and some talkers were "L-like."

We hypothesized that native speakers' "R-L" decisions take into account momentary changes in talker characteristics. Various results indicate that when subjects listen to native-language stimuli, a change in talker requires them to normalize the characteristics of the new talker (i.e., to "tune" to the new talker; see [4, 1]). Japanese subjects, in contrast, develop criteria based on the range of cues to /r/ and /l/ they hear in a session. That is, they adapt to the range of available cues rather than the relative trade-offs in cues that accompany talker changes

(as suggested by Nusbaum and Lee [5]).

2 Experiment: Effects of Proportion

Our explanations of previous results have rested on the assumption that Japanese subjects expect to hear equal numbers of /r/s and /l/s in an experimental setting, and choose criteria for "R-L" decisions that divide the range of cues they are hearing into two categories of roughly equal numbers. This assumption has been tested with synthetic speech [6] but not with natural speech. We tested this assumption by manipulating the proportion of /r/s and /l/s subjects heard. We predicted, in accordance with our previous explanations, that Japanese subjects' rates of "R" response would not be affected by the proportion manipulation. We also manipulated talker condition (stimuli from single talkers *blocked* vs. stimuli from multiple talkers *mixed*), in order to replicate earlier results. To ensure that any effects observed were not artifacts of talker intelligibility or the task, we conducted the same experiment with native speakers of American English. In order to avoid ceiling effects in accuracy, the stimuli native speakers of American English heard were degraded. We predicted that the proportion manipulation *would* affect American English subjects' rates of "R" response.

2.1 Method
2.1.1 Subjects

Two groups of subjects participated in the experiment. A group of 15 native speakers of Japanese who had never lived abroad and had no history of hearing or speech disorders, and a group of 5 native speakers of American English (AE subjects) with no history of hearing or speech disorders.

2.1.2 Stimuli

The stimuli consisted of sixty minimal pairs of monosyllabic English words which contrast /r/ and /l/ in initial position (e.g., "rate" vs. "late"). We selected the productions of three native speakers of American English (AE) used by Magnuson and Yamada [3]; two of the talkers appeared "R-like" in their experiments, and one appeared "L-like." These talkers were from a set previously recorded and used by Logan et al. [2]. The stimuli were the same for both sets of subjects, except that we added noise to the stimuli used for the American English subjects to prevent their accuracy from staying at ceiling level. We added noise by randomly choosing 15% of the samples of each stimulus, and changing the sign of those samples. This results in "noisy," degraded stimuli. Pilot studies showed that the 15% degradation consistently results in accuracy levels of 70% to 80% with AE subjects.

2.1.3 Procedure

We manipulated talker condition (*blocked* versus *mixed*) and proportion of /r/s and /l/s. There were three levels of the proportion manipulation: *equal* (50% /r/s, 50% /l/s), *more R* (70% /r/s, 30% /l/s), and *more L* (70% /l/s, 30% /r/s). Thus, each subject participated in six blocks: blocked-equal, mixed-equal, blocked-*more R*, mixed-*more R*, blocked-*more L*, and mixed-*more L*.

In previous experiments some subjects have taken more than 30 trials to adjust to a change in range. To

prevent confounding adjustments to talkers with adjustments to proportions, each block was made up of two sequences of stimuli, parts A and B (see Figure 1). For the blocked condition, two lists of 30 stimuli were randomly generated from the set of 60 stimuli for each talker. For both lists, the proportion of /r/s and /l/s was based on the proportion condition. The first 90 stimuli of each blocked condition consisted of 30 stimuli from an "R-like" talker, followed by 30 stimuli from the "L-like" talker, followed by 30 stimuli from the other "R-like" talker.

The sequence of the last 90 stimuli (part B) was similar – 30 "R-like," 30 "L-like", 30 "R-like" – but was generated from a second list of 30 stimuli for each talker. Mixed blocks were generated by randomizing the order of part A of the appropriate blocked-talker block, and following that sequence by part B of the appropriate blocked-talker block in random order. Analyses were based only on the last 90 responses in each block.
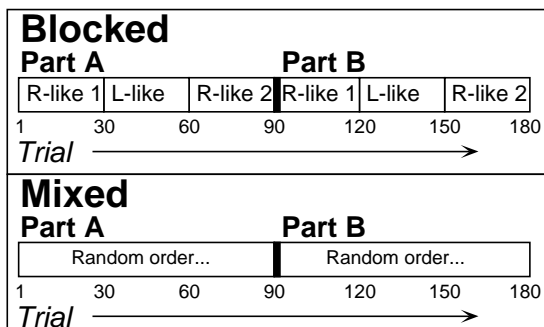

Figure 1. Control of talker condition and proportion manipulations. In the blocked condition, stimuli were blocked by talker. In the mixed condition, the stimuli from the two blocked sequences were randomly ordered.

2.2   Results

An analysis of the rate of "R" response (how often subjects responded "R" on all trials – including correct and incorrect responses) reveals a significant effect of *talker* ($F(2,28)=12.169$, p=.0002). Confirming previous results, subjects responded "R" most often to the talkers we had identified as "R-like" in previous experiments, and less often to the "L-like" talker. There was also a significant interaction of *talker* and *talker condition* (*blocked* or *mixed*). The pattern of greater rates of "R" response for "R-like" talkers was augmented by the talker condition manipulation (see Figure 2).

An analysis of variance in rate of "R" response of AE subjects shows that, unlike Japanese subjects, AE subjects' rate of "R" response changed significantly in response to changes in proportion of /r/s and /l/s ($F(2,8)=8.012$, p=.0123; see Figure 3).

An analysis of variance comparing rate of "R" response for Japanese and AE subjects shows a significant effect of *native language* ($F(1,18)=4.729$, p=.043), and a significant interaction of *proportion* and *native language* ($F(2,36)=10.164$, p=.0003; see Figure 3).

3   Discussion

The significant difference in Japanese subjects' rate of "R" response from blocked to mixed conditions confirms previous findings (e.g., [3]), that Japanese subjects are not sensitive to variability due to talker changes when making "R-L" decisions. In addition, Japanese subjects' rate of "R" response did not change when the proportion of /r/ and /l/ stimuli was changed, although the AE subjects' rate of "R" response did. These results confirm previous findings that Japanese subjects adapt
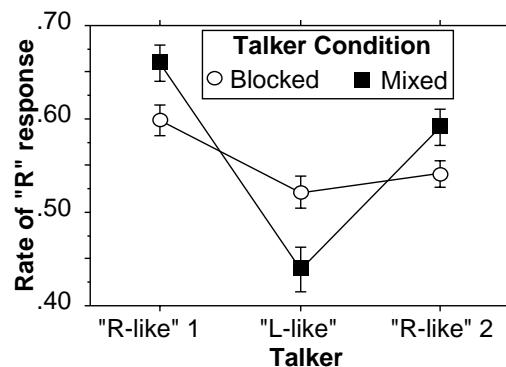

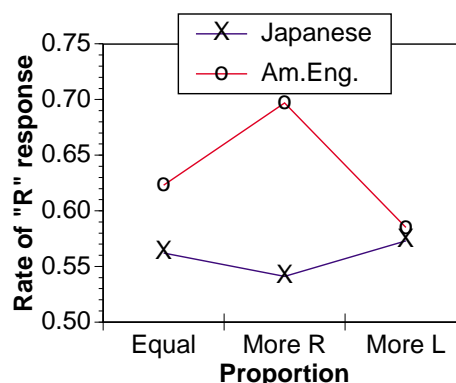Figure 2. Interaction of talker and talker condition in Japanese subjects' rate of "R" response.

Figure 3. Effect of proportion on rate of "R" response.

to changes in the range of cues available for "R-L" decisions by setting criteria that divide the range of cues into two categories, rather than adapting to changes in talker characteristics [6, 3] (e.g., normalizing talker differences). Indeed, the demonstration that Japanese subjects are not sensitive to momentary stimulus-intrinsic (talker changes) or longer-term stimulus-extrinsic (proportion of /r/s and /l/s) changes in cues to /r/ and /l/ strongly indicates that Japanese subjects are developing novel mappings from the range of cues they hear in each session to "R" and "L" categories.

REFERENCES

[1] 加藤和美、 筧一彦 (1988). 音声知覚における話者への適応性の検討. 日本音響学会誌, 44, 180-186.

[2] Logan, J. S., Lively, S. E., and Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. J. Acoust. Soc. Am., 89, 874-886.

[3] Magnuson, J. S. and Yamada, R. A. (1994). Talker variability and the identification of American English /r/ and /l/ by Japanese listeners. J. Acoust. Soc. Am., 95, 2872.

[4] Nusbaum, H. C., and Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, Y. Sagisaka, and E. Vatikiotis-Bateson (Eds.), Speech Perception, Speech Production, and Linguistic Structure, pp. 113-134. Tokyo: OHM.

[5] Nusbaum, H. C., and Lee, L. (1992). Learning to hear phonetic information. In Y. Tohkura, Y. Sagisaka, and E. Vatikiotis-Bateson (Eds.), Speech Perception, Speech Production, and Linguistic Structure, pp. 265-173. Tokyo: OHM.

[6] Yamada, R. A., and Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. Perc. & Psychophysics, 52, 376-392.