日本人大学院生における
LとRの音声弁別能力の実態調査

Discrimination Ability Between L and R Speech Sounds among Japanese Graduate Students

秋元頼孝 / 三宅圭音 / 鈴木理玖

AKIMOTO Yoritaka / MIYAKE Keito / SUZUKI Riku

キーワード
子音の弁別、日本人大学院生、信号検出理論

Keywords
consonant discrimination, Japanese graduate students, signal detection theory

本研究では、日本人大学院生を対象に、日本人が苦手とする英語の /r/ と /l/ の弁別能力について調査した。31 名が 5 つのブロックからなる /ra/ と /la/ の弁別実験に参加した。第 1 のブロックでは、120 個の /la/ と 30 個の /ra/ の音声を 1 秒の間隔でランダムな順序で提示し、実験参加者は /ra/ が聴覚提示された時にキーを押すという課題を行った。第 2 ブロックでは、/ra/ と /la/ の音声刺激の数を逆にしたうえで、/la/ が聴覚提示された時にキーを押すという課題を行った。第 3 ブロックはキー押しを行わない学習ブロックであり、/la/ と /ra/ を 3 回ずつ交互に合計 150 回、聴覚と視覚で同時提示した。第 4、第 5 ブロックでは、それぞれ第 1、第 2 ブロックと同じ課題を行った。実験の結果、良好な弁別成績を示したのは実験参加者の約 10%であった。一方、約半分の実験参加者の弁別成績はチャンスレベル程度であった。また、第 3 ブロックで聴覚・視覚同時提示を経験した後に、/ra/ の検出課題における反応時間が有意に短くなり、また弁別能力が向上する傾向も認められた。本研究は、英語のリスニング教育において有用な基礎データを提供するものである。

Introduction

It is widely known that the Japanese have difficulty distinguishing between the /r/ and /l/ speech sounds. For example, Yamada and Tohkura (1992) reported the rate of correct identification of naturally spoken English words containing /r/ or /l/ consonants was 65.7% on average in Japanese who had no experience of living in a foreign country. Simultaneously, great individual differences were reported in this study; approximately 40% of the participants' identification accuracy was less than 60%, 30% of the participants' identification accuracy was within 60% to 70%, and 20% of the participants' identification accuracy was within 70% to 80%. Further, approximately 10% of the participants' identification accuracy was more than 80%. These data are valuable because they report not only average performance but also the distribution of individual performance, which is especially essential in the context of English listening education. Similarly, Ueda et al. (2007) reported that the rate of correct identification of English words containing /r/ or /l/ was 67.2% on average, with approximately 20% of the participants' identification accuracy being less than 60%, 50% of the participants' identification accuracy was within 60% to 70%, 20% of the participants' identification accuracy was within 70% to 80%, and 10% of the participants' identification accuracy was more than 80%.

Ueda et al. (2007) also reported that 15 days (each 2 hours) identification training of /r/ - /l/ improved the performance of eight Japanese participants whose initial identification accuracy was within 50% to 70%. On the final day, six of them reached over 90% correct, one about 85% correct, and one 75% correct. A previous study also reported directional asymmetry in the discrimination between /r/ and /l/. For example, Kuhl et al. (2006) used a conditioned heart-turn procedure, a widely used technique for the assessment of infants' ability to perceive speech, to examine their ability to discriminate consonants. They found that detection of the stimulus change from /ra/ to /la/ is more difficult than from /la/ to /ra/ regardless of age (i.e., 6 or 12 months) or native language (Japanese or English) in infants. Yamada and Tohkura (1992) and Ueda et al. (2007) used the same number of /r/ and /l/ stimuli, and the distribution of individual discrimination performance by the direction of the stimulus change was not reported.

This study surveyed Japanese graduate students' ability to discriminate between /r/ and /l/ sounds. Specifically, we examined the discrimination ability of /ra/ and /la/ speech sounds with considering the change in direction, using the signal detection theory (Green & Swets, 1966). This could clarify the percentage of Japanese people who could distinguish between /r/ and /l/ sounds and the extent to which they could do so in each direction. Furthermore, we examined the effect of a short learning experience (less than 3 min). We hypothesized that a short-term learning experience might increase discrimination performance in /ra/ detection from /la/ sounds but not in /la/ detection from /ra/ sounds, because the latter was shown to be challenging in a previous study (Kuhl et al., 2006).

1. Methods

1.1 Participant

The experiment was conducted as a class assignment for graduate students enrolled in "Advanced Experimental Psychology" at the Nagaoka University of Technology. All were native Japanese, and all or most had no experience of studying abroad, although this was not individually confirmed. Since the experiment was conducted as part of the class assignment, the participants could not refuse to participate in the experiment. However, they were free to decline using their data for purposes other than class. On the day of the experiment, 29 men and two women in their early 20s attended the class, and none of them declined to use their data for purposes other than the class.

After the experiment, the log files were distributed to the students for data analysis, which was also a part of the class assignment and a final report for the course credit. Given that the first author was the lecturer in charge of the class, we determined the analysis method before the students' report submission, so that reading students' reports would not lead to plagiarism. Consequently, no reports used signal detection theory to accomplish the same analysis as this study.

1.2 Apparatus

The experiment was conducted using students' laptops running Windows 10. Students who did not have a Windows 10 laptop used other students' laptops for the experiment. Further, students used their earphones or those prepared by the experimenter. No specific instructions were provided regarding the volume level, and the participants could adjust the volume as they liked.

1.3 Stimuli

A total of 10 speech sounds of /ra/ and /la/, produced by 10 female native English speakers, were used in this experiment. Since all authors were Japanese and were not confident in auditory discrimination of /ra/ and /la/ sounds, we used Praat, software for sound analysis, to visually check the shape of the third formant trajectory of the stimuli, given that /r/ and /l/ have different characteristics in the shape of the third formant trajectory (Gordon et al., 2001).

1.4 Procedure

The experiment was conducted simultaneously in the classroom, divided into the first half (those who had a Windows laptop) and the second half (those who borrowed other students' laptops) during one class period (90 min). First, the participants were instructed that the experimental task was to distinguish between /r/ and /l/ and to remain quietly in the classroom when they were not experimenting. Next, the experimental program created by PsychoPy (Peirce et al., 2019) was distributed to the participants, and they wore earphones and ran the program independently. No practice trials were conducted because we wanted to gather data without practice effect. Additionally, the participants were not informed that the ratios of /ra/ and /la/ speech sounds were not the same and changed depending on the experimental block.

The experimental procedure is shown in Figure 1. In the first block of the experiment, the instruction that "Press Enter for Ra, but not for La" appeared on the screen, and 2 s later, playback of the auditory stimuli began while keeping the instruction on the screen. The stimuli of the first block comprised 30 /ra/ and 120 /la/ speech sounds, which were played in a pseudo-random order with 1 s stimulus onset asynchrony (i.e., from the start of stimulus presentation to the start of the next stimulus presentation at intervals of 1 s). After the first block was completed, the instruction that "Press Enter for La, but not for Ra" appeared on the screen for 2 s, and then the second block started. The stimuli in the second block comprised 30 /la/ and 120 /ra/ speech sounds in pseudo-random order. After the second block was finished, the instruction of the third block that "Listen to La and Ra sounds. Letters appearing on the screen correspond to audio. No key-presses required." appeared on the screen. Two seconds later, /ra/ and /la/ were both auditory and visually presented alternately three times each, for 150 times. As was the instruction, visually presented letters on the screen corresponded to auditory presented sounds. The fourth and fifth

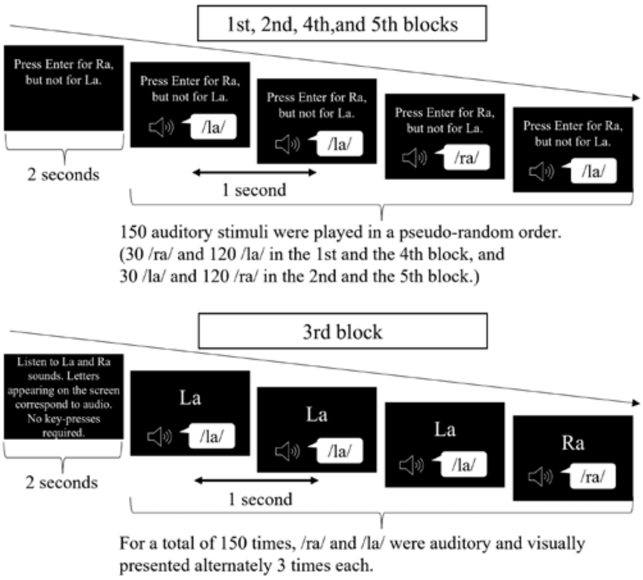

Figure 1. Procedure of the experiment

blocks were identical to the first and second blocks. The experiment took approximately 13 min.

1.5 Signal detection theory

In evaluating the discrimination ability, it is necessary to evaluate both the performance of (1) correctly pressing the key when the target speech sounds (i.e., hit) and (2) correctly not pressing the key when the non-target speech sounds (i.e., correct rejection) in an integrated manner. Therefore, we evaluated each participant's discrimination ability using d' (d-prime), a measure of signal detection theory (Green & Swets, 1966) that distinguishes between the real sensitivity of participants and their potential response biases. The d' indicates the distance between the distribution of trials that include the signal and the distribution of trials that do not include the signal (i.e., noise trials) concerning a psychological quantity corresponding to the intensity of the signal. In summary, a larger d' value indicates a higher discriminative power because there is a greater distance between the distributions of the signal and noise trials. If d' is close to zero, there is no discriminative power because the distributions of the signal and noise trials almost overlap. If d' is negative and not close to zero, signal and noise trials distribution are distinguished but reversed. In this case, the signal and noise trials are perceptually discriminated, but the required responses are the opposite. Specifically, participants found the signal from the noise trials and did not find signals from the signal trials.

2. Results

The frequencies of the occurrence of hit, miss, correct rejection, and false alarm for each participant in each block are presented in Table 1. As shown in Table 1, we found several negative d' values that were not close to zero, especially in the second and fourth blocks. This is strange but could happen if some participants were unaware that the detection target had changed from /ra/ to /la/ in the second or fifth experimental blocks. Therefore, we asked the

Table 1. Values of d' for each participant for each block. The number of hit, miss, correct rejection, and false alarm shows the number of frequency of their occurrence. |d'| indicates mean absolute d' value in 1st, 2nd, 4th, and 5th experimental blocks. IDs are assigned in order of mean performance

	[mean]	[1st block] Detection of /ra/ (30 /ra/, 120 /la/)					[2nd block] Detection of /la/ (30 /la/, 120 /ra/)				
ID	|d'|	Hit	Miss	Correct Rejection	False Alarm	d'	Hit	Miss	Correct Rejection	False Alarm	d'
1	2.87	27	3	116	4	3.12	24	6	107	13	2.08
2	2.73	27	3	117	3	3.24	22	8	109	11	1.95
3	2.16	26	4	78	42	1.50	24	6	116	4	2.68
4	1.59	21	9	102	18	1.56	26	4	70	50	1.32
5	1.57	27	3	88	32	1.90	8	22	47	73	-0.90
6	1.52	24	6	107	13	2.08	3	27	68	52	-1.11
7	1.51	21	9	78	42	0.91	18	12	110	10	1.64
8	1.42	18	12	82	38	0.73	26	4	85	35	1.66
9	1.35	24	6	63	57	0.90	26	4	92	28	1.84
10	1.29	22	8	100	20	1.59	8	22	33	87	-1.22
11	1.08	13	17	108	12	1.11	4	26	75	45	-0.79
12	0.88	12	18	56	64	-0.34	17	13	101	19	1.17
13	0.83	15	15	109	11	1.33	2	28	103	17	-0.43
14	0.79	17	13	78	42	0.55	23	7	76	44	1.07
15	0.74	17	13	58	62	0.13	14	16	62	58	-0.04
16	0.69	11	19	84	36	0.18	9	21	94	26	0.26
17	0.58	19	11	60	60	0.34	16	14	48	72	-0.17
18	0.50	18	12	89	31	0.90	18	12	67	53	0.40
19	0.49	14	16	93	27	0.67	14	16	84	36	0.44
20	0.46	16	14	54	66	-0.04	15	15	55	65	-0.10
21	0.43	19	11	56	64	0.26	17	13	51	69	-0.02
22	0.43	20	10	55	65	0.33	19	11	53	67	0.19
23	0.42	13	17	60	60	-0.17	11	19	39	81	-0.79
24	0.40	18	12	74	46	0.55	11	19	71	49	-0.11
25	0.36	12	18	79	41	0.15	11	19	81	39	0.11
26	0.34	16	14	47	73	-0.19	19	11	42	78	-0.04
27	0.34	19	11	61	59	0.36	15	15	59	61	-0.02
28	0.32	11	19	53	67	-0.49	12	18	83	37	0.25
29	0.26	11	19	87	33	0.26	16	14	87	33	0.68
30	0.25	8	22	63	57	-0.56	16	14	67	53	0.23
31	0.12	14	16	67	53	0.06	14	16	73	47	0.19
mean	0.93	17.7	12.3	78.1	41.9	0.74	15.4	14.6	74.5	45.5	0.40
	[mean]	[4th block] Detection of /ra/ (30 /ra/, 120 /la/)					[5th block] Detection of /la/ (30 /la/, 120 /ra/)				
ID	|d'|	Hit	Miss	Correct Rejection	False Alarm	d'	Hit	Miss	Correct Rejection	False Alarm	d'
1	2.87	28	2	119	1	3.90	24	6	113	7	2.41
2	2.73	25	5	117	3	2.93	24	6	117	3	2.80
3	2.16	25	5	105	15	2.12	21	9	116	4	2.36
4	1.59	27	3	92	28	2.01	27	3	68	52	1.45
5	1.57	25	5	107	13	2.20	3	27	61	59	-1.26
6	1.52	24	6	100	20	1.81	23	7	77	43	1.09
7	1.51	27	3	83	37	1.78	22	8	103	17	1.70
8	1.42	23	7	101	19	1.73	24	6	92	28	1.57
9	1.35	27	3	62	58	1.32	25	5	78	42	1.35
10	1.29	23	7	81	39	1.18	9	21	32	88	-1.15
11	1.08	14	16	112	8	1.42	12	18	107	13	0.98
12	0.88	23	7	71	49	0.96	22	8	80	40	1.05
13	0.83	13	17	106	14	1.02	7	23	70	50	-0.52
14	0.79	20	10	76	44	0.77	23	7	61	59	0.75
15	0.74	28	2	80	40	1.93	23	7	66	54	0.85
16	0.69	23	7	90	30	1.40	8	22	47	73	-0.90
17	0.58	23	7	60	60	0.73	7	23	44	76	-1.07
18	0.50	14	16	71	49	0.15	18	12	74	46	0.55
19	0.49	18	12	71	49	0.49	10	20	95	25	0.38
20	0.46	19	11	78	42	0.73	22	8	77	43	0.99
21	0.43	23	7	61	59	0.75	23	7	59	61	0.71
22	0.43	23	7	39	81	0.27	23	7	70	50	0.94
23	0.42	11	19	56	64	-0.42	13	17	54	66	-0.29
24	0.40	16	14	82	38	0.56	14	16	82	38	0.39
25	0.36	16	14	82	38	0.56	17	13	81	39	0.62
26	0.34	25	5	60	60	0.97	19	11	37	83	-0.16
27	0.34	21	9	68	52	0.69	16	14	69	51	0.27
28	0.32	10	20	63	57	-0.37	17	13	61	59	0.19
29	0.26	12	18	76	44	0.09	18	12	49	71	0.02
30	0.25	18	12	50	70	0.04	17	13	44	76	-0.17
31	0.12	19	11	49	71	0.11	13	17	73	47	0.11
mean	0.93	20.7	9.3	79.6	40.4	1.09	17.5	12.5	72.8	47.2	0.58

participants about this possibility the week after the experiment. Consequently, we found that four participants were unaware of the change in the instruction throughout (ID 5, ID 10, ID 13, and ID 17 seem to be these participants from the pattern of the d' values). Five participants were not aware of the change in the instruction immediately but noticed it during the experiment (ID 6 and ID 11 seem to be these participants from the pattern of the d' values). The reasons for this were that some of them performed the experiment with their eyes closed to concentrate on the auditory stimuli, and the other performed the experiment with their eyes open but were too concentrating to notice the change in the instruction on the screen. It was impossible to confirm which participants had done so because of time constraints and the priority of the progress of the class. However, because a negative d' value not close to zero also indicates perceptual discrimination success, as described in the previous section, we decided to use absolute d' values for discrimination performance to address this problem. Note that discrimination performances in some of the participants who noticed the instruction change during the second or the fifth blocks became underestimated because the absolute value of the d' becomes smaller in this case. The absolute d' values for each participant for each block plotted in receiver operating characteristic space are presented in Figure 2. The mean absolute d' values in the first, second, fourth, and fifth blocks were 0.74, 0.40, 1.09, and 0.58, respectively.

To examine the directional asymmetry and the effect of the short-term learning effect, we conducted a two-way analysis of variance (ANOVA) with absolute d' as the dependent variable and detection target (/ra/ vs. /la/) and short-term learning experience (before the third block vs. after the third block) as within-participants variables. The results revealed a significant main effect of the detection target ($F(1, 30) = 10.32$, $p < 0.01$, partial $\eta^2 = 0.256$), showing that absolute d' was higher in the /ra/ detection condition than in the /la/

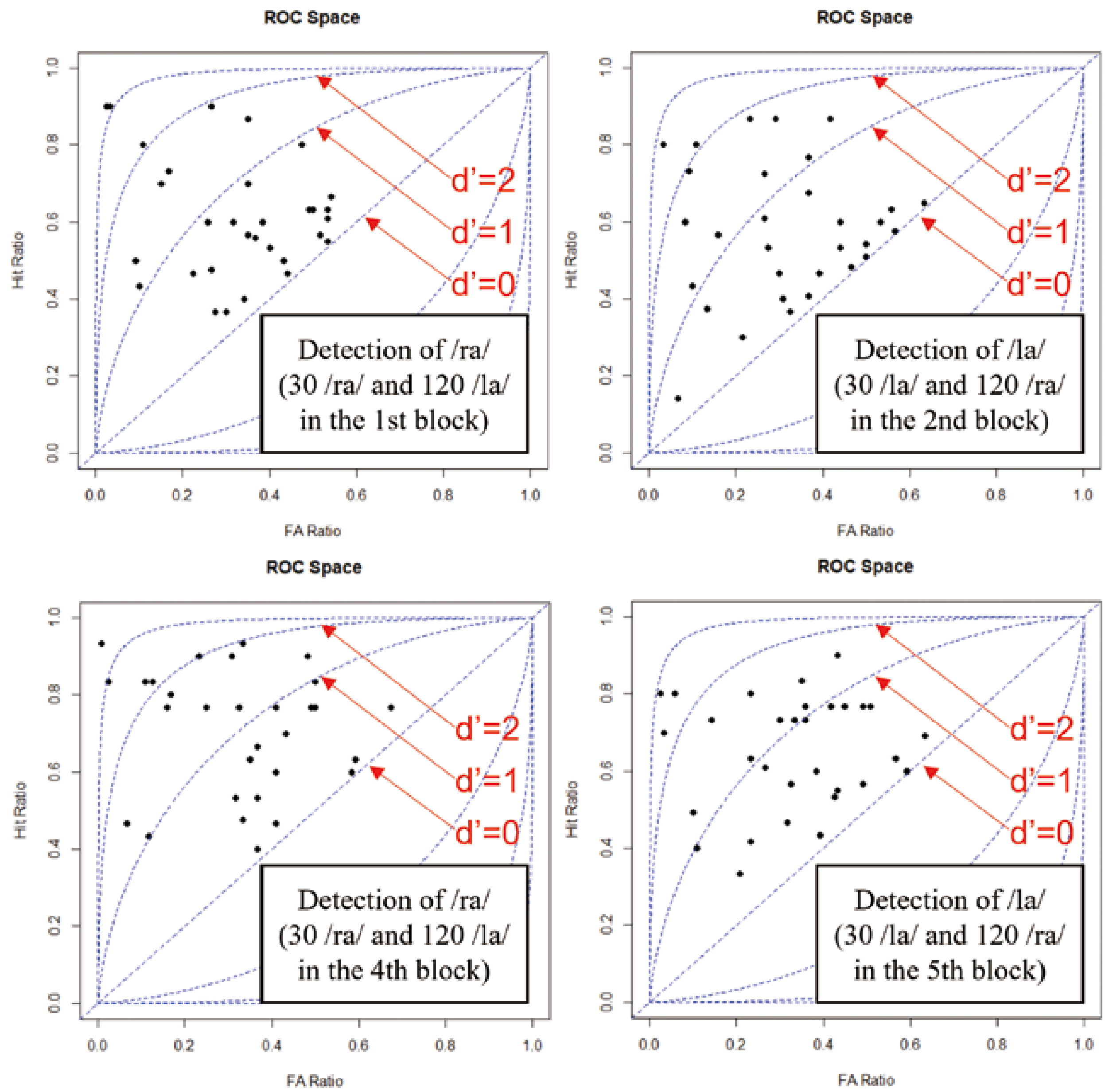

Figure 2. Absolute d' values for each participant for each block plotted in Receiver Operating Characteristic (ROC) Space. The x and y axes indicate false alarm (FA) ratio and hit ratio, respectively. A plot in the upper left corner indicates good performance, while a plot on the diagonal from lower left to upper right indicates no discriminative power

Table 2. Correlations between the absolute d' values in each block

	1st block	2nd block	4th block	5th block
1st block	-	.69 **	.80 **	.75 **
2nd block		-	.70 **	.82 **
4th block			-	.86 **
5th block				-

(** p < .01)

detection condition, and a marginally significant main effect of the short-term learning experience (F (1, 30) = 10.32, $p < 0.10$, partial $\eta^2 = 0.092$), suggesting that absolute d' increased by the short-term learning experience. The correlations of the absolute d' values in each block were also provided for readers' interests (Table 2).

Given that the reaction time might reflect the difficulty of the detection, we also conducted a three-way ANOVA with reaction time as the dependent variable and correctness of response (correct vs. incorrect), detection target (/ra/ vs. /la/), and learning experience (before the third block vs. after the third block) as within-participants variables. The rationale was that, although participants were not instructed to press the key quickly, they should do so because the stimuli were presented quickly (i.e., with 1 s stimulus onset asynchrony). The results showed a significant interaction between detection target and learning experience (F (1, 30) = 6.15, $p < 0.05$). Follow-up analysis revealed a significant effect of learning experience in the /ra/ detection condition ($p < 0.05$) and a significant effect of detection target after the learning experience ($p < 0.01$). The reaction times in the first, second, fourth, and fifth blocks were 0.52, 0.52, 0.50, and 0.53 seconds, respectively.

3. Discussion

We surveyed the discrimination ability between /r/ and /l/ sounds in Japanese graduate students to determine the number of Japanese who could distinguish between L and R sounds and to what extent. Although some participants performed the task inappropriately in the second and fifth experimental blocks, we could evaluate the discrimination ability with no serious problems by using absolute d' values.

Although greater absolute d' values indicate a higher discriminative power, we need to examine the data closely to determine what value should be considered excellent. First, we checked the minimum value of d' in the first and fourth experimental blocks, where all participants performed the task appropriately because the detection target was the same as that shown at the beginning of the experiment; thus, overlooking it did not matter in the fourth experimental block. Consequently, we found that it was -0.56, and we could assume that at least this magnitude was within the range of what could happen by chance. Based on this, we categorized the participants whose mean absolute d' values were under 0.56 (ID = 18–31, Table 1) as the worst performance group, whose discrimination ability was as good as the chance level (n = 14). We then checked these participants' maximum absolute d' value, which might be considered the maximum absolute d' value by chance. Since we found that it was 0.99, we categorized the participants whose mean absolute d' values were under 1.00 (ID = 12–17, Table 1) as the second-worst group (n = 6). We then examined the high-performing side. We found that three participants showed outstanding performance, with mean d' values over 2.16 (ID = 1–3, Table 1). The next highest performing participant's mean d' value was 1.59. Therefore, we decided to adopt more than 2.00 in mean absolute d' values as a criterion for the best performance group (n = 3). The remaining participants (ID = 4–11; Table 1) were categorized into the intermediate group (n = 8). This distribution ratio was similar to that reported by Yamada and Tohkura (1992) and Ueda et al. (2007), indicating that our criteria were reasonable and comparable to those of previous studies.

The results also showed that detecting /la/ sounds from /ra/ sounds were more difficult than vice versa. This directional asymmetry is consistent with a previous study on infants' consonant discrimination tasks (Kuhl et al., 2006). We also found that short-term learning experiences of simultaneous auditory and visual presentation could improve discrimination performance to some degree, although it was only marginally significant. Additionally, we found that the reaction time was shortened after the learning experience in detecting /ra/ from the /la/ condition, which was generally consistent with the discrimination performance results.

In conclusion, this study revealed that approximately 10% of Japanese graduate students showed excellent performance in discriminating between /ra/ and /la/ sounds, and 50% of Japanese graduate students showed poor performance as good as the chance level. This study also provided the nature of the distribution of individual discrimination performance in Japanese graduate students based on signal detection theory in each direction of the stimulus change, which is especially useful for English listening education.

References

Gordon, P. C., Keyes, L., & Yung, Y. F.: Ability in Perceiving Nonnative Contrasts: Performance on Natural and Synthetic Speech Stimuli, Perception & Psychophysics, 63 (4), pp. 746–758, 2011

Green, D. M., & Swets, J. A.: Signal Detection Theory and Psychophysics, Wiley, pp. 1969–2012, 1966.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P.: Infants Show a Facilitation Effect for Native Language Phonetic Perception between 6 and 12 Months, Developmental Science, 9 (2), F13-F21, 2006.

Kusanagi, K.: Theory of Signal Detection from 1/0 Correct/Incorrect Data: Self-Developed Function of R. Kusanagi's Research Log. https://kusanagi.hatenablog.jp/entry/2015/08/13/235533, August 13, 2015. (accessed 25 August 2022)

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K.: PsychoPy2: Experiments in Behavior Made Easy. Behavior Research Methods, 51 (1), pp. 195–203, 2019.

Ueda, K., Akahane-Yamada, R., Komaki, R., & Adachi, T.: Identification of English /r/ and /l/ in Noise: The Effects of Baseline Performance, Acoustical Science and Technology, 28 (4), pp. 251–259, 2007.

Yamada, R. A., & Tohkura, Y. I.: The Effects of Experimental Variables on the Perception of American English /r/ and/ l/ by Japanese Listeners, Perception & Psychophysics, 52 (4), pp. 376–392, 1992.