Effects of language background on perception of whole speech and chirp-only non-speech stimuli: The case of /ra/-/la/ contrast*

☆Kanako Tomaru, Takayuki Arai (Sophia Univ.)

1 Introduction

1.1 Purpose of the study

It is widely known that native language (L1) affects second language (L2) perception. One of the well-known examples is perception of the English /r/-/l/ contrast by native speakers of Japanese (for example, [1] and [2], among others). Japanese-speaking listeners are said to have trouble with perceptually differentiating /r/ from /l/ because the Japanese phoneme inventory does not include these segments as separate phonemes. The study of Miyawaki *et al.* (Miyawaki75) [2] is one of the researches that concerned with perception of the /r/-/l/ contrast by English- and Japanese-speaking listeners.

Miyawaki75, investigated effects of language background on speech perception. In their experiment, native speakers of English and those of Japanese perceptually discriminated synthetic /ra/-/la/ syllables along a continuum. They demonstrated that English-speaking listeners showed a categorical perception [3] whereas Japanese-speaking listeners did not. Such results of Miyawaki75 suggested that only the L1 consonantal contrast was perceived categorically.

In addition to the experiment above, they also had the same groups of listeners to discriminate chirps extracted from the /ra/-/la/ continuum. The interesting finding was that perception of the non-speech chirp stimuli, i.e. transitions of the third formant (F3), was not influenced by the listeners' native language. That is, effects of language background were shown to be limited to speech perception.

In the present research, we attempt to replicate the former findings of Miyawaki75: 1) a categorical perception is observed for L1, not L2, speech perception (finding 1), and 2) perception of non-speech stimuli is not affected by listeners' language background (finding 2). For our experiment, we employed similar, but slightly different conditions.

1.2 Conditions for the present study

In Miyawaki75, two types of stimuli were used: 1) a synthetic /ra/-/la/ continuum (whole speech) with 15 steps, and 2) an F3 chirp continuum (chirp-only non-speech) with 13 steps. The /ra/-/la/ continuum consisted of the first three formants, i.e. F1, F2, and F3. For the experiment, native speakers of English (EN) and those of Japanese (JP) were recruited. Firstly, EN and JP listeners participated in an oddity discrimination (ODD) task; both whole speech and chirp-only non-speech stimuli were used. In the ODD task, listeners were to judge which one of the triad differed from the others. The stimuli to be compared differed in three steps in each continuum. In addition, an identification (ID) task was also conducted for the whole speech stimuli; only EN listeners participated in the task. Prior to the main experimental sessions, Miyawaki75 gave listeners thorough practice sessions to familiarize them with the stimuli. See Table 1 for details.

In the present study, following Miyawaki75, we had a synthetic /ra/-/la/ continuum as whole speech stimuli and an F3 chirp continuum as chirp-only non-speech stimuli. However, details of the materials and experimental procedures were slightly different from those of Miyawaki75. In the present research, the whole speech stimuli consisted of the first five formants instead of three. The number of steps was also different: ten steps for the whole speech continuum, nine steps for the chirp-only non-speech continuum. For the experimental task, we employed an AXB discrimination (AXB) task for EN and JP listeners, and an ID task for EN listeners. Different groups of EN listeners were recruited

*音声(音節)と非音声(チャープ音)の知覚における母語の影響:/ra/-/la/の対比の場合, 渡丸嘉菜子, 荒井隆行(上智大・理工).

Table 1. The difference between experimental settings of Miyawaki *et al.* and those of the present study.

	Miyawaki75				Present study				
	Whole speech		Chirp-only non-speech		Whole speech			Chirp-only non-speech	
Number of steps	15		13		10			9	
Formants involved	F1-F3		F3		F1-F5			F3	
Listeners	EN	JP	EN	JP	EN-1	EN-2	JP	EN-2	JP
Participated task(s)	ID & ODD	ODD	ODD	ODD	ID	AXB	AXB	AXB	AXB
Practice	Yes, thorough stimuli familiarization				No, only task familiarization				

for each task (EN-1 and EN-2 in Table 1). In the AXB task, listeners were to judge whether the second sound (X) best matched to the first (A) or to the third (B). Moreover, our practice session was short, and it was only for task familiarization, rather than stimuli familiarization. Table 1 indicates the difference between experimental settings of Miyawaki75 and those of the present research.

Through the perceptual experiment, it was suggested that listeners' language background affected perception of speech, but not non-speech, as suggested by Miyawaki75, regardless of differences in the experimental conditions.

2 Experiment

2.1 Materials

2.1.1 Whole speech continuum

A series of /ra/-/la/ continuum was created using cascade-formant software synthesizer designed by Klatt and Klatt [4]. The /ra/-/la/ syllables were created based on a male speaker's utterance from the TIMIT corpus [5]: the speaker ID was MKAM0. For the synthesis, we obtained formant frequency values of the speaker from a vowel [ʌ] in a word, "pronunciation" of a sentence "Clear pronunciation is appreciated" (the sentence ID was sx236). Table 2 indicates the speaker's first three formant frequencies of the selected part of the vowel [ʌ], averaged over time. We used these values as steady state values of the vowel /a/ in the /ra/-/la/ syllables.

Following MacKain *et al.* [1], we calculated onset frequencies of F1, F2, and F3 (*F1s*, *F2s*, and *F3s* in Fig. 1). For F3, we also calculated the values at the inflection at 135 ms. Figure 1 provides schematic representation of trajectories of the first five formants. Following Miyawaki75, only *F3s* and the value at the inflection varied in

Table 2. The speaker's frequencies of the first, second and third formants of the selected part of the vowel /ʌ/, averaged over time.

Formant	Value (Hz)
F1	670
F2	1357
F3	2788

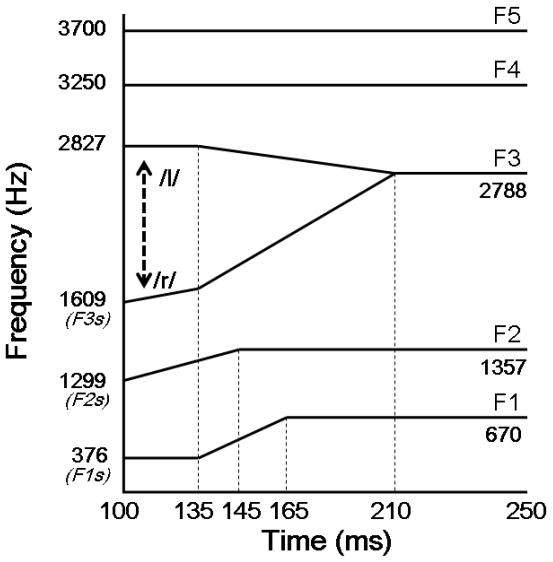

Fig. 1. Schematic representation of formant trajectories from F1 to F5 without the rising or the falling period.

nearly equal ten steps from /ra/ (Speech-Step1) to /la/ (Speech-Step10), i.e. from 1609 Hz to 2827 Hz for *F3s* and from 1717 Hz to 2827 Hz for the inflection. *Fs1* and *Fs2* were fixed throughout the continuum. Default values of the synthesizer were used for F4 (3250 Hz) and F5 (3700 Hz), and the values were fixed throughout the continuum.

All synthesized syllables were 350 ms-long with 100-ms rising and falling periods of amplitude. Figure 1 shows the 250-ms long period without the rising or the falling period. Amplitude in the rising and the falling periods was changed linearly in the decibel scale from 0 dB at 0 ms to 60 dB at 100 ms, and from 60 dB at 250 ms to 0 dB at 350 ms by using the parameter, "amplitude of voicing (AV),"

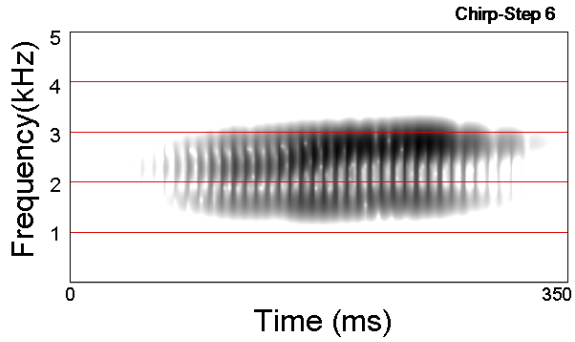

Fig. 2. The spectrogram of a chirp, Chirp-Step6.

of the synthesizer. During the rising period, all formats had the onset values which are illustrated in Figure 1. For the experiment, additional 100 ms silent periods were added before and after each of the synthesized syllables.

In addition to the first three formants, the speakers' F0 contour of the "pronunciation" part of the utterance was approximated, and reflected in the synthesized syllables. Digital outputs from the synthesizer (10-kHz sampling rate and 16-bit resolution) were converted to 16-kHz sampling rate and 16-bit resolution.

2.1.2. Chirp-only non-speech continuum

To create the chirp-only non-speech continuum, we filtered nine synthesized /ra/-/la/ syllables (Speech-Step 1 through Speech-Step 9). Filtering was done by using "Filter (pass Hann band)" provided by the Praat software [6]. The center of the lower transition band was 1590 Hz, and that of the upper transition band was 2800 Hz. The transition bandwidth was 20 Hz. The frequency range of F3 transition of the nine syllables were 1609 Hz to 2707 Hz. We had F3 chirp continuum differed in nine steps (Chirp-Step 1 to Chirp-Step 9) as non-speech stimuli. All chirps had upward transition. Figure 2 shows an example.

2.2 Listeners

2.2.1. Native speakers of English

Two groups of EN listeners were recruited. One group consisted of nine people (6 males, 3 females) participated in the AXB task of the whole speech continuum, and the chirp-only non-speech continuum. The other group consisted of two people (1 male, and 1 female) participated in the ID task of the speech continuum. None reported any known hearing problems.

2.2.2. Native speakers of Japanese

One group that consisted of 14 JP listeners with normal hearing (4 males, 10 females) was recruited to participate in the AXB task of the whole speech and the chirp-only non-speech continua. The ID task was not conducted for the JP listeners following Miyawaki75.

2.3 Procedure

All sessions were carried out using Praat software [6]. Stimuli were presented diotically via Sennheiser HDA 200 headphones at participants' comfortable listening level.

2.3.1 ID task

Only EN listeners participated in the ID task. The listeners were instructed to choose what they've heard is either "ra" or "la". Participants heard four repetitions of each of the ten stimuli, all of which were presented randomly to participants. Thus, participants made total of 40 judgments for each continuum (4 repetitions×10 stimuli). Listeners took a practice session to be familiarized with the procedure.

2.3.2 AXB task

EN and JP listeners participated in the AXB task. Both whole speech and chirp-only non-speech stimuli were used for this task. Both types of stimuli were paired such that each pair (AB) differed by two steps in the continuum, i.e. Speech-Step 1－3, Speech-Step 2 － 4, …, Speech-Step 8 － 10; Chirp-Step 1－3, Chirp-Step 2－4, …, Chirp-Step 8 －10. Note that speech were always paired with speech, and non-speech were always paired with non-speech. Listeners were instructed to judge if the second syllable, or sound (X), matches to the first (A), or to the third (B), and to guess if necessary. Paired stimuli were arranged into four permutations (AAB, ABB, BAA, and BBA). There were three repetitions for each presentation, so listeners made 12 judgments for each pair. Thus, this makes total of 96 judgments for one session, i.e. syllable or chirp (8 pairs × 4 presentations × 3 repetitions = 96 judgments). AXB presentations were made randomly within each session. A short practice session was held to familiarize listeners with procedures. All listeners had chirp discrimination task first.

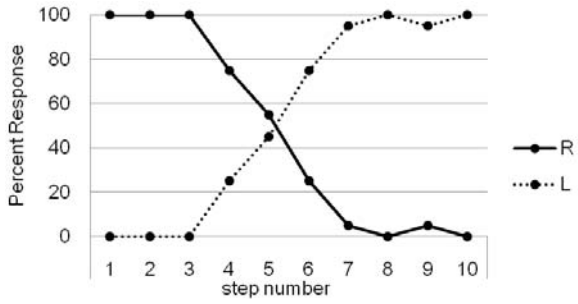

Fig. 3. Average percent response (%) of /ra/ (solid line) and /la/ (dotted line) for English listeners in the ID task.

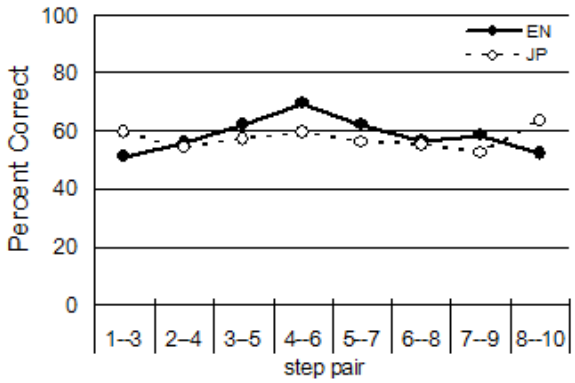

Fig. 4. Average correct rate (%) of discrimination of syllables for EN (solid line) and JP listeners (dashed line).

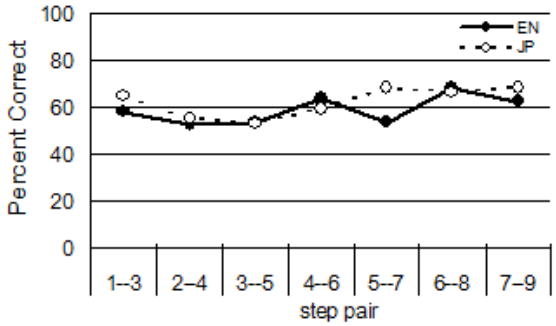

Fig. 5. Average correct rate (%) of discrimination of F3 chirp for EN (solid line) and JP listeners (dashed line).

3 Results

3.1 ID task

Percent responses were averaged with EN listeners (Fig. 3). The categorical boundary was shown to locate at Step 5, where the /ra/ responses dropped from 75% at Step 4 to 55%.

3.2 AXB task: whole speech

Percent correct was averaged within EN and JP listeners (Fig. 4). The discrimination function of JP listeners looks flat throughout the continuum, whereas that of EN listeners indicates a peak at the pair 4－6, which crosses the categorical boundary indicated from the ID task (cross-category pair). Nevertheless, the difference between the two groups at the pair 4－6 was not significant: t (20) = 1.84, p = .081.

3.3 AXB task: chirp-only non-speech

Figure 5 indicates averaged correct rate at each step for EN and JP listeners. Discrimination function for each group does not seem to diverge greatly from each other. At the pair 5－7, we see a gap between correct rate for EN listeners and that for JP listeners. However, the difference was not significant: t (20) = -.425, p = .645.

4 Conclusion

The present study recruited similar, but different experimental settings to replicate the findings of Miyawaki75. Results of the AXB discrimination of whole speech stimuli revealed that the continuum was perceived categorically by EN listeners, but not by JP listeners: a discrimination peak was observed at the cross-category pair only for EN listeners (finding 1). However, the difference between EN and JP listeners at the cross-category pair was not significant. This may be because the two-step comparison was too difficult even for EN listeners that the accuracy stayed lower than we expected. In addition, discrimination functions for the chirp-only stimuli did not differ depending on listeners' language background (finding 2). Overall, our experiments supported the previous findings that only speech perception is affected by listeners' language background.

References

[1] MacKain *et al.*, Appl. Pshycholing., 2 (4), 369-390, 1981.

[2] Miyawaki *et al.*, Percept. Psychopys., 18 (5), 331-340, 1975.

[3] Liberman *et al.*, J. Exp. Psycho., 61 (5), 379-388, 1961.

[4] Klatt and Klatt, JASA, 87 (2), 820-857, 1990.

[5] Zue *et al.*, Speech Comm., 9 (4), 351-356, 1990.

[6] Boersma & Weenink, Glot International, 5 (9), 341-345.