ICA 2019 AACHEN

PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Discrimination of mono-syllables in sentence context: the case of Japanese listeners' perception of /ba/-/da/ continuum

Kanako TOMARU[1],[2]

[1] Mejiro University, Japan

[2] Sophia University, Japan

ABSTRACT

It is widely known that native speech sounds are perceived according to the phonetic categories of the native language: the process is called categorical perception. The hypothesis of categorical perception assumes that two stimuli are perceived to be different only when these are identified as different. Therefore, the best discrimination performance (discrimination peak) is observed when the paired stimuli are members of different categories. Traditionally, categorical perception has been tested using monosyllables presented in isolation. However, my recent research has shown that the discrimination peak is not observed when the mono-syllabic stimuli are embedded in a sentence. This recent research investigated the categorical perception in a sentence using an AXB paradigm with a stimulus interval of 300 ms: The relatively short interval may be one of the factors that suppressed the discrimination peak. To clarify this point, the current study conducted a perceptual experiment on Japanese listeners using an AXB paradigm with a stimulus interval of 1 s. The results suggest that for the perception of a Japanese /ba/-/da/ continuum by Japanese listeners, the length of the interval has only a small impact.

Keywords: Categorical perception, Sentence context, Stimulus interval

1. INTRODUCTION

The process of speech perception involves two common abilities, namely the ability to identify phonemes and the ability to discriminate one phoneme from another. The combination of these abilities leads to categorical perception (1). Strictly speaking, categorical perception assumes that listeners can discriminate two sounds only when these sounds are identified as different phonemes. For example, when native speakers of English perceive a sound continuum that gradually changes from English /r/ to /l/ in ten steps, they divide the continuous stimuli into two categories, namely /r/ (Step 1 through Step 4) and /l/ (Step 6 through Step 10), placing a boundary in the middle (Step 5). When discriminating stimuli along the continuum, listeners use the category to make discrimination judgments. In other words, the discrimination performance for two sounds can be predicted from listeners' identification results. Therefore, the hypothesis of categorical perception predicts that two stimuli from the same category will be heard as being the same. In contrast, the best discrimination performance (discrimination peak (2)) is observed when the paired stimuli are members of different phonemic categories (cross-boundary pairs). The fundamental principle that discrimination performance can be predicted from identification results can be adopted for cross-language speech perception. For native speakers of Japanese, English /r/ and /l/ are heard as the single Japanese phoneme /ɾ/ (e.g., (3,4)); therefore, there appears to be no categorical boundary along the /ra/-/la/ continuum. Thus, no discrimination peak is observed when native speakers of Japanese perceptually discriminate these English phonemes.

Traditionally, categorical perception is observed through identification and discrimination experiments using monosyllabic stimuli presented in isolation. However, recent research has shown that the characteristics of categorical perception, especially a discrimination peak, disappear when the monosyllabic stimuli are embedded in a sentence. For instance, Tomaru and Arai (5,6) reported that when native speakers of English discriminate English /ra/ syllables from /la/ syllables, discrimination

[1] ka-tomaru@sophia.ac.jp

[2] k.tomaru@mejiro.ac.jp

(a) (b)

performance does not increase at the cross-boundary pairs. For their experiments, synthetic stimuli that continuously change from English /ra/ to /la/ were synthesized. The experimental tasks included identification and discrimination tests under three sound conditions: (i) in isolation (5,6), (ii) between pure tones (6), and (iii) within a sentence (5,6). Their experiments revealed that discrimination performance under conditions (i) and (ii) was consistent with the traditional prediction of categorical perception, i.e., discrimination was highly accurate when comparing stimulus pairs from different phoneme categories. However, under condition (iii), discrimination accuracy did not increase even for a comparison of cross-boundary pairs. Subsequent research conducted by Tomaru and Arai (7) found the same tendency for Japanese listeners' perception of voiced stops, i.e., /ba/ and /da/.

The purpose of the current research is to replicate the preceding findings (i.e., the characteristics of categorical perception are not found under the sentence condition) using an AXB paradigm with a longer inter-stimulus interval (ISI). In the previous research (5–7), listeners' discrimination performance was assessed using an AXB paradigm with an ISI of 300 ms, which is adequate for examining the categorical perception of monosyllables in isolation (8). However, for sentence perception, it is possible that a short ISI increases the perceptual burden. Under the sentence condition, English syllables, for example /ra/-/la/, are embedded into an English sentence, for example "Clear /ra/ (or /la/) is appreciated" (5, 6). During an AXB discrimination task, listeners hear the sentence three times consecutively, and judge whether X matches A or B. Although speech perception is a relatively fast automatic process (9), it may be difficult for listeners to complete two steps, namely identify the stimuli in question and compare them, during a short period of time. If discrimination performance is negatively affected by a short ISI, then a longer ISI should yield a discrimination peak. The current research investigates Japanese listeners' categorical perception of /ba/-/da/ syllables under two conditions: (i) in isolation and (ii) within a sentence. The results of a perceptual experiment suggest that the length of an ISI does not affect listeners' discrimination performance under the sentence condition.

2. Method

2.1 Stimuli

In the current study, a nine-step continuum of /ba/-/da/ syllables is employed. Under the isolation condition, the continuum of syllables was presented to participants in isolation. Under the sentence condition, the same syllables were embedded into a sentence, e.g., *sorekara /ba/ ga aruto omoimasu* (/ba/ was embedded), "*And, I think there is /ba/.*" The target syllable, the embedded /ba/ in this example, changes in nine steps along the continuum.

The nine-step /ba/-/da/ continuum originally created for previous experiments was used here (7). The continuum was created using the cascade-formant synthesizer designed by Klatt and Klatt (10,11). The parameter values for creating syllables were decided on the basis of a /ba/ utterance by a male native speaker of Japanese： *sorekara /ba/ ga aruto omoimasu* "And I think that there is /ba/." The mean values of formants 1 through 5 were obtained from the token and used as the parameter values. In addition to the formant frequencies, the bandwidth of each formant, extra tilt of the voicing spectrum, and quotient of vocal-fold opening were manipulated on the synthesizer to imitate the speaker. To create a continuum, the formant trajectories of F1 and F2 were varied in nine steps from canonical /ba/ (Step 1) to /da/ (Step 9). The values of the fundamental frequency (F0) were decided so that the created syllables would naturally fit the original sentence when presented within a sentence under the sentence condition. Please refer to (7) for further information on stimulus synthesis.

2.2 Participants

Twenty native speakers of Japanese (7 males and 13 females) ranging in age from 19 to 32 years (mean age = 24.1 years) participated in the experiment. The participants reported no hearing difficulties at the time of the experiment.

2.3 Experimental Methods

The participants performed identification and discrimination tasks under the isolation condition and the sentence condition. Each participant had a familiarization session before experimental sessions. All sessions were conducted using Praat software (12) and headphones connected to a computer via a USB-connected audio interface.

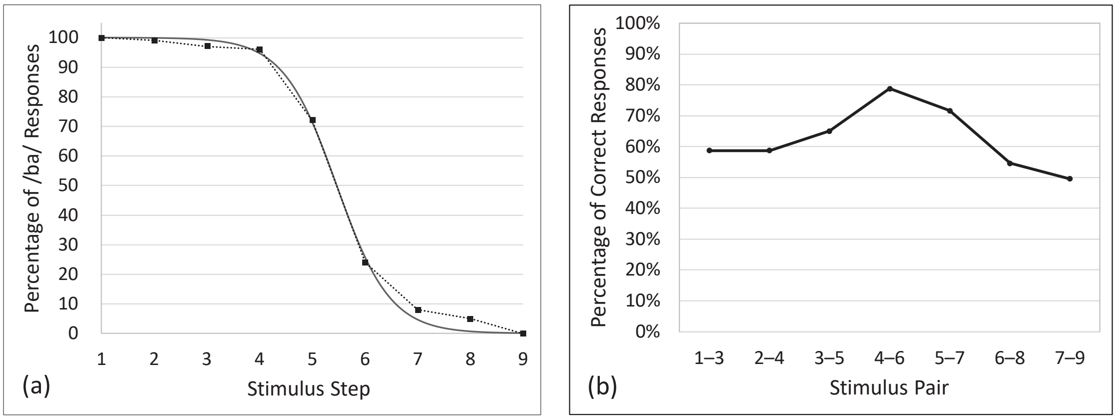

Figure 1 – Results of identification (a) and discrimination (b) tasks under the isolation condition.

2.3.1. Familiarization

Before the experimental sessions, participants were exposed to the synthesized syllables to become familiarized with artificial sounds. First, the participants heard five repetitions of the edge stimuli, i.e., Step 1 and Step 9. Then, the continuum was presented twice. The first time, the participants heard the continuum from Step 1 to Step 9 in ascending order; the second time, they heard the same continuum in descending order. During the familiarization session, the participants were asked to make self-judgments about whether the stimuli they heard were /ba/ or /da/. The participants were able to adjust the sound to a comfortable listening level during this session. They were asked not to change the volume once the main experiments had begun.

2.3.2. Identification Task

The two-alternative forced-choice method (2AFC) was utilized. Participants were asked to judge whether the syllable they heard was “ば /ba/ ” or “だ /da/” by clicking on the appropriate button in the computer interface. The participants performed the task under the isolation condition and the sentence condition in separate sessions. For both conditions, each stimulus (syllable or sentence) along the continuum was repeated five times in random order (9 stimuli × 5 repetitions = 45 judgments). Practice sessions preceded experimental sessions.

2.3.3. Discrimination Task

The AXB paradigm, where participants were asked to judge whether X matches A or B, was employed. As done in the identification session, the participants performed the AXB discrimination task under the isolation condition as a separate session from that for the sentence condition. For both conditions, the stimuli were paired such that each pair differed by two steps along the continuum (e.g., Step 1 and Step 3). In addition, the edge stimuli (i.e., Step 1 and Step 9), were paired and presented to participants as fillers. Each pair was repeated three times in random order (8 pairs × 3 repetitions). For both conditions, the ISI was 1 s. Practice sessions were conducted before experimental sessions.

Under the isolation condition, participants were instructed to compare syllables and to judge whether the second syllable (X) matched the first (A) or third (B) syllable. Similarly, under the sentence condition, the participants were told to concentrate on the syllables in the middle of a sentence, and to compare three sentences. Then, they were asked to judge whether the second sentence matched the first or third sentence.

3. RESULTS

3.1 Isolation Condition

3.1.1. Identification Task

The identification results obtained under the isolation condition are summarized in Fig. 1-(a). The participants’ responses are averaged and shown by the dotted line. The solid line shows the percentage of /ba/ responses fitted using the following logistic model.

$$y = 100/1 + e^{a(x-b)} \qquad (1)$$

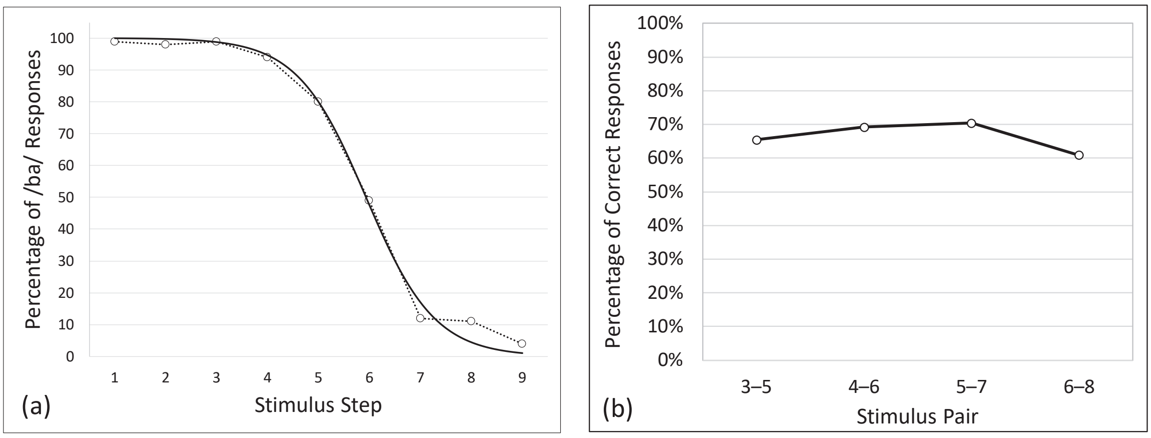

Figure 2 – Results of the identification (a) and discrimination (b) tasks under the Sentence Condition.

,where y is the percentage of /ba/ responses and x is the step number on the continuum. The parameters a and b represent the slope of the curve and the 50% crossover point, respectively. In this case, the slope of the curve is 2.0 and the 50% crossover point is 5.5. Based on the identification results, we can predict a discrimination peak at the pair Step 4 and Step 6, which crosses the category boundary.

3.1.2. Discrimination Task

The percentage of correct responses was calculated for each listener. Figure 1-(b) shows the average percentage of correct responses. The identification results indicate that the category boundary is at 5.5. Therefore, the discrimination peak is predicted for the pairs 4–6, 5–7. Assuming that these pairs are cross-boundary pairs, the stimulus pairs were divided into three groups: (i) the within /ba/ pairs, i.e., pairs 1–3, 2–4, and 3–5, (ii) the cross-boundary pairs, i.e., pairs 4–6 and 5–7, (iii) the within /da/ pairs, i.e., pairs 6–8 and 7–9. A one-way analysis of variance (ANOVA) revealed a main effect of the group ($F(2, 137) = 23.46$, $p < 0.001$). A post hoc multiple comparison with Bonferroni correction found a significant difference between the groups: the percentage of correct responses for the within /ba/ pairs was different from that for the cross-boundary pairs ($p < 0.01$). Similarly, the percentage of correct responses for the within /da/ pairs was different from that for the cross-boundary pairs ($p < 0.01$). In addition, the difference between the within /ba/ group and the within /da/ group was significant ($p < 0.05$). The results indicate that the observed characteristics of categorical perception are as predicted under the isolation condition.

3.2 Sentence Condition

3.2.1. Identification Task

The identification results obtained under the sentence condition are shown in Fig. 2-(a). The dotted line indicates the average responses of all participants. The average responses were fitted using the model in Eq. (1). The solid line indicates the fitted function. Under the sentence condition, the slope of the curve is 1.5 and the 50% crossover point is 5.9. Thus, if the characteristics of categorical perception are to be observed under this condition, a discrimination peak should appear at the stimulus pairs 4–6 and 5–7.

3.2.2. Discrimination Task

The discrimination results are summarized in Fig. 2-(b). Under this condition, more than seven listeners' responses for some within-category pairs, i.e., the pairs 1–3, 2–4 and 7–9, were below a chance level, ranging from 17% to 42%. This may suggest that listeners intentionally gave responses that were opposite to those they were expected to provide. Thus, the results for these pairs were eliminated from the analysis.

As shown in Fig. 2-(b), no discrimination peak was observed under the sentence condition. As done in the analysis under the isolation condition, the stimulus pairs were classified into three groups: (i) the within /ba/ group, i.e., the pair 3–5, (ii) the cross-boundary group, i.e., the pairs 4–6, 5–7, and (iii) the within /da/ group, i.e., the pair 6–8. A one-way ANOVA found no main effect of the group ($F(2, 77) = 2.07$, $p = 0.13$). No significant difference between the groups was observed.

4. Discussion and Conclusion

The purpose of the research was to investigate whether a discrimination peak appears under the sentence condition with an ISI of 1 s. The results of the perceptual experiments show that even with a 1-s ISI, the sentence context hinders discrimination performance. That is, the lack of a discrimination peak under the sentence condition is not the result of a perceptual burden caused by a short ISI.

The hypothesis of categorical perception predicts a discrimination peak because it assumes that listeners can discriminate two sounds only when they have different phonemic labels. If an AXB paradigm with a short ISI, e.g., 300 ms (5–7), causes an excessive burden for listeners, preventing them from completing the labeling process and making discrimination judgments at the same time, then the same paradigm with a relatively long ISI (e.g., 1 s), should reduce this burden; as a result, discrimination results should reflect labeling results. However, this was not the case. Therefore, the results of the current experiment support the idea that listeners do not use labeling information during sentence discrimination. This is consistent with previous studies (5–7). Further, the lack of a discrimination peak suggests that the ability to discriminate sounds based on phonemic categories is limited to the perception of sounds in small units (e.g., syllables). When perceiving sentences, listeners either do not discriminate sounds at all or rely on information other than phonemic labels.

In addition, it is worth mentioning that listeners' discrimination performance is unstable under the sentence condition compared to that under the isolation condition. The results of the current experiment showed that the percentage of correct responses did not reach a chance level in many cases for within-category pairs. A previous study (7) reported similar results, where several listeners' percentage of correct responses for the pairs 1–3 and 2–4 was remarkably low, ranging from 8% to 42%. These outcomes may imply that sentence discrimination and syllable discrimination should be treated differently. Remaining issues will be investigated in future research.

ACKNOWLEDGEMENTS

I would like to express sincere gratitude to Professor Takayuki Arai and the members of the Arai Laboratory at Sophia University for providing helpful comments and feedback on the current experiment.

This research was supported by a Japan Society for the Promotion of Science (JSPS) KAKENHI grant (JP26884051, JP18K12378).

REFERENCES

1. Liberman AM, Harris KS, Hoffman HS, Griffith BC. The discrimination of speech sounds within and across phoneme boundaries. J Exp Psychol. 1957;54:358-368.
2. Howell P, Rosen S. Natural auditory sensitivities as universal determiners of phonemic contrasts. Linguistics. 1983;21:205-236.
3. Guinon SG, Flege JE, Akahane-Yamada R, Pruitt JC. An investigation of current models of second language perception: The case of Japanese adults' perception of English consonants. J Acoust Soc Am. 2000;107(5):2711-2724.
4. MacKain KS, Best CT, Strange W. Categorical perception of English /r/ and /l/ by Japanese bilinguals. Appl Psycholinguist. 1981;2:369-390.
5. Tomaru K, Arai T. Perception of /ra/-/la/ contrast in different contexts: mono-syllable vs. sentence. Proc Meet Acoust; 2-7 June 2013; Montreal, Canada 2013. p. 1-9.
6. Tomaru K, Arai T. Discrimination of /ra/-/la/ speech continuum by native speakers of English under nonisolated conditions. Acoust Sci Tech. 2014;35(5):251-259.
7. Tomaru K, Arai T. Role of labeling mediation in speech perception: Evidence from a voiced stop continuum perceived in different surrounding sound contexts. Acoust Sci Tech. 2016;37(6):303-314.
8. Pisoni DB. Auditory and phonetic memory codes in the discrimination of consonants and vowels. Percept Psychophys. 1973;13:235-260.
9. Johnson K, Ralston JV. Automaticity in speech perception: Some speech/nonspeech comparisons. Phonetica. 1994;51:195-209.
10. Klatt DH. The new MIT speech VAX computer facility. In: Research Laboratory of Electronics ed. Speech Communication Group Working Papers IV. MIT: Cambridge; 1984. p. 73-82.
11. Klatt DH, Klatt LC. Analysis, synthesis, and perception of voice quality variation among female and male talkers. J Acoust Soc Am. 1990;87:820-857.
12. Boersma P, Weenink D. Praat, a system for doing phonetics by computer. Glot Int. 2001;5(9):341-347.