Japanese listeners' sensitivity to the difference between fricatives and affricates goes down as surrounding context gets linguistic*

☆ Kanako Tomaru, Takayuki Arai（Sophia University）

1 Introduction

Former studies suggest not only foreign sounds, but also native sounds with weak phonological relations are often difficult to discriminate [1-3]. For example, allophones of the same phoneme are difficult to discriminate for native speakers of the language in question. It is difficult for English listeners to discriminate [ɾ] from [d], which are in allophonic relation in English phonology, e.g. both [ɹaɪɾɚ] and [ɹaɪdɚ] are acceptable to mean "rider" [4]. Although the number of studies that directly deal with the impact of allophony on perception is limited, we can find reports that imply the effects of allophony on perception [5-7], and it seems valid to suggest that allophones are difficult to discriminate.

Now, are allophones always difficult to discriminate? Listeners are less sensitive to allophonic difference than they are to phonemic contrast at least in natural speech with linguistic structure, e.g. naturally produced syllables, but what would happen if allophones are presented in non- natural-speech context? The purpose of the current study is to reveal if allophones are difficult to discriminate when they are not part of natural speech. To reveal this point, the current experiment recruited Japanese listeners to discriminate allophonic variants of Japanese /z/, i.e. [z] and [dz], presented in different types of non-natural-speech contexts: natural-speech-like monosyllable, disyllable, phrase, and sentence contexts, and a pure tone context. The results of the experiment show that Japanese listeners' sensitivity to allophonic difference is high in the pure tone context, but it reduces when allophones are part of linguistic structure of natural-speech-like context. The results suggest that allophones become difficult to discriminate when they are not part of natural speech, as long as the context forms linguistic structure.

2 Experiment

2.1 Purpose of the experiment

The purpose of the current study is to reveal if allophones are difficult to discriminate when they are not part of natural speech. In order to accomplish the goal, the current experiment recruited Japanese listeners to discriminate a voiced fricative [z] and a voiced affricate [dz], which are allophonic variants of /z/ in Japanese, presented in different types of non-natural-speech contexts: natural-speech-like monosyllable, disyllable, phrase, and sentence contexts, and a pure tone context. Henceforth, the former four contexts are called "linguistic contexts," and the pure tone context is called "non-linguistic context". The details of stimuli are reviewed in the following section.

If allophonic difference is always difficult to detect for Japanese listeners despite presented context, the listeners' sensitivity to the [z]-[dz] difference would be low regardless of the type of context, i.e. linguistic or non-linguistic. If the sensitivity to allophonic difference varies, it is highly probable that the listeners would be insensitive to the difference in the linguistic contexts because allophonic difference is imperceptible in linguistic structure in natural speech. In addition, if allophones are not even heard as allophones in the non-linguistic context, listeners' sensitivity to the difference in such context is predicted to be high.

In addition to allophonic variants of /z/, a voiceless fricative [s] and a voiceless affricate [ts], which are not in allophonic relations, are employed for comparison, in order to confirm that the results obtained for the target sounds

*日本語話者の摩擦音・破擦音の識別力は言語コンテキスト内で低下する, 渡丸　嘉菜子, 荒井 隆行 (上智大).

reflect the impact of allophony.

2.2 Recording

The stimulus materials were constructed from recordings produced by the first author, a female native speaker of Japanese. The speaker recorded a sentence /kinoɯ jaCVga kieta/ where CV was substituted by 24 syllables with eight consonants and three vowels: 1) [za, zi, zɯ], 2) [dza, dzi, dzɯ], 3) [sa, si, sɯ], 4) [tsa, tsi, tsɯ], 5) [ta, ti, tɯ], 6) [da, di, dɯ], 7) [ʃa, ʃi, ʃɯ], and 8) [ʒa, ʒi, ʒɯ]. The nonsense word /jaCV/ was read as an unaccented word. The recorded sentences were used to create stimuli. The recording was carried out in a sound-proof studio onto a digital recorder at the sampling rate of 48 kHz (Marantz, PDM660).

2.3 Stimuli preparation

The current research attempts to reveal Japanese listeners' sensitivity to allophonic difference presented in different types of surrounding context, i.e. the linguistic contexts and the non-linguistic context, by using the same-different discrimination test. A linguistic context consists of natural-speech-like sequences of sounds, which has a linguistic structure such as a monosyllable, a disyllable, a phrase, or a sentence. Therefore, to present target consonants in the linguistic context means to present them as part of linguistic structure of such. For example, a target consonant [z] is in the linguistic context when it is presented as a syllable [za], or as a sentence [kinoɯ jazaga kieta]. On the other hand, the non-linguistic context consists of a pure tone surrounding the target allophones. Thus, a target consonant is in the non-linguistic context when it is presented with a pure tone immediately proceeds and follows it.

During the experiment, listeners were to judge if the paired stimuli sound the same or different, and they were to focus only on the difference between the target consonants. Therefore, stimuli pairs of the current experiment must form minimal pairs in a strict way differing only in the consonant in question.

The first step to create the linguistic context stimulus pairs was to create minimally different sentences using recorded materials. To create such sentences, the author first took a sample of [a], [i], and [ɯ] out of the recorded materials, and put them into the place of V in one of the recorded sentences, /kinoɯ jaCVga kieta/, which actually was one token of [kinoɯ jadɯga kieta]. This created three new sentences varying only in V as depicted in (1), where newly replaced segments are marked with underlines. The new sentences were used as templates for stimuli creation for the target allophones, [z] and [dz].

(1) Template sentences
   i. [kinoɯ jad a ga kieta]
   ii. [kinoɯ jad i ga kieta]
   iii. [kinoɯ jad ɯ ga kieta]

Then, the author took one sample of target fricative [z] and one sample of target affricate [dz] out of recorded sentences to put into the place of [d] in the template sentences. Now we have additional two sentences for each of the three sentences varying in the vowel as depicted in (2)-(4), where newly replaced segments are marked with underlines. These sentences make minimal pairs within the same vowel group. The sentences were used as base sentences to create stimuli.

(2) Base sentences with a following vowel [a]
   i. [kinoɯ ja z aga kieta]
   ii. [kinoɯ ja dz aga kieta]

(3) Base sentences with a following vowel [i]
   i. [kinoɯ ja z iga kieta]
   ii. [kinoɯ ja dz iga kieta

(4) Base sentences with a following vowel [ɯ]
   i. [kinoɯ ja z ɯga kieta]
   ii. [kinoɯ ja dz ɯga kieta]

In the process of stimuli creation, a 1 kHz pure tone was used to cover up the sides of a base sentence, leaving a target consonant alone (the context "C"), or with linguistic structures such as a monosyllable (the context "CV"), a disyllable (the context "CVCV"), a phrase (the context "CVCVga"), or a sentence (the context

"Sentence"). The pure tone covered up both sides of a base sentence just like covering both sides of a window with curtains, leaving a crack for a peek (Fig. 1). Covered up sentences remained with a target consonant with linguistic structures, i.e. the contexts CV, CVCV, CVCVga, and Sentence, were used as linguistic context stimuli (Fig.1-a, for example). Covered up sentences remained only with a target consonant appearing in the middle, i.e. the context C, were used as non-linguistic context stimuli (Fig.1-b). This makes five contexts for one consonant as shown in (5), where blacken parts illustrate pure tones. The examples of the stimuli are illustrated in Figure 1. We prepared each type of stimuli for both of the target consonants, [z] and [dz], for every vowel.

(5) Contexts used in the experiment

C:	e.g.	z
CV:	e.g.	za
CVCV:	e.g.	jaza
CVCVga:	e.g.	jazaga
Sentence:	e.g.	[kinoɯ jazaga kieta]

Control stimuli with [s] and [ts] were created using the identical method as above, except that the template sentence was created based on a token of [kinoɯ jasɯga kieta]. All stimuli were created using Praat software [8]. The signal-to-pure-tone ratio was -13.3 dB. The calculated values were based on the average RMS of all base sentences and the RMS of a pure tone.

2.4 Experimental procedure

Twenty Japanese listeners were recruited to be participants. The experimental task was to judge if paired sounds A and B were the same or not. Target allophones and control consonants were always paired with their partners, i.e. [z] was always paired with [dz], and [s], [ts]. And, type of contexts for the partners always matched, e.g. if [z] was presented in the context CV, its partner [dz] was also presented in the context CV, and so on. Therefore, the paired sounds always formed a minimal pair, differing only in C in /jaCV/. Listeners heard two types of the same presentation (AA, BB) and two types of the different presentation (AB, BA). All listeners heard one repetition of each presentation for each pair. The listeners made the same-different judgment 120 times (2 consonant pairs × 3 vowels × 5 contexts × 4 presentations × 1 repetition = 120 judgments).

2.5 Results

Collected responses were converted to d-prime (d') scores in order to avoid any response bias [9]. Perfectly accurate responses receive the perfect score of d', which is 2.77 in this case. The d' score of zero indicates no sensitivity to the difference, where Hit Rate equals False Alarm.

Figures 2 and 3 show the results for the target

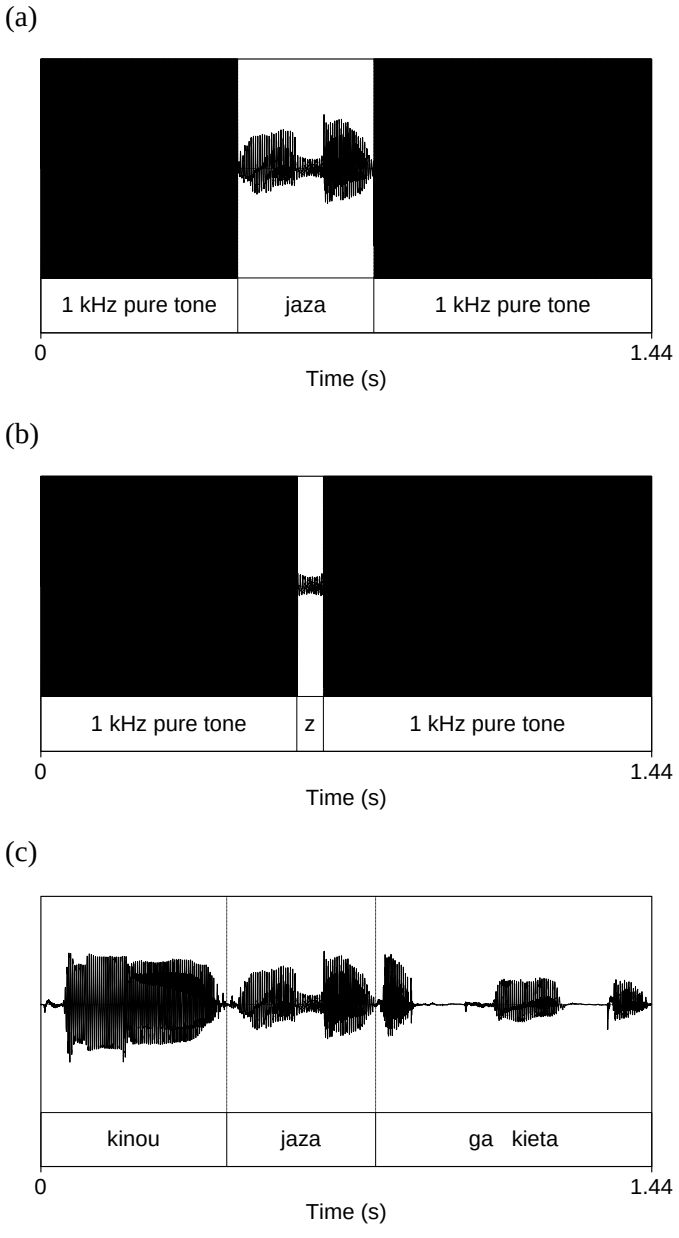

Fig. 1 An example of CV stimulus (a) and a C stimulus (b) created from the base sentence (c), which also serves as a sentence stimulus itself.

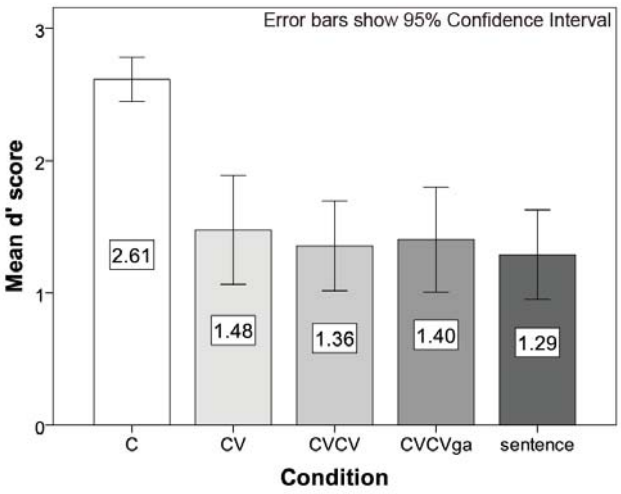

Fig. 2 Japanese listeners' sensitivity to the difference between [z] and [dz] shown in d'.

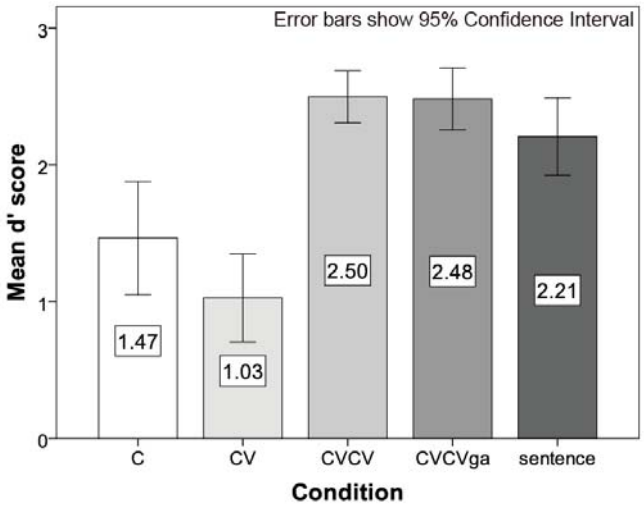

Fig. 3 Japanese listeners' sensitivity to the difference between [s] and [ts] shown in d'.

pairs and the control pairs, respectively. Since the performance varies depending on the pair, i.e. target or control, it is confirmed that the results for the target sounds reflect the impact of allophony.

As seen in Figure 2, listeners were accurate in discriminating the target consonants when they were presented in the context C. However, the sensitivity dramatically dropped when the consonants were presented in the other contexts. A one way ANOVA with repeated measures showed the main effect of the contexts (F (2.96, 54.39) = 18.27, p < .001), and the listeners' performance in the context C was significantly better than the other contexts (p < .001).

On the other hand, the listeners' sensitivity to the difference between the control consonants was rather low in the C and the CV contexts, while the sensitivity was constantly high in the rest of the contexts. A one way ANOVA with repeated measures revealed the significant main effect of the contexts (F (2.26, 42.90) = 26.60, p < .001). The significant difference was found between the C and the CVCV contexts (p = .001), the C and the CVCVga (p = .003), the CV and the CVCV (p < .001), the CV and CVCVga (p < .001), the CV and the Sentence (p < .001); the significance between the C and the Sentence contexts was only marginal (p = .069).

3 Conclusion

The experiment revealed that Japanese listeners' sensitivity to the allophonic difference is high in the non-linguistic context, but it reduces in the linguistic contexts. The results suggest that sounds must be part of a linguistic structure to be heard as allophones, but context do not have to be absolutely natural.

Acknowledgments

This study was partially supported by Sophia University Open Research Center from MEXT.

References

[1] Goto, Neuropsychologia, 9 (3), 317-323, 1971.

[2] Hume & Johnson, In Hume & Johnson (Eds.) *The Role of Speech Perception in Phonology*, Academic Press, 2001.

[3] Hume & Johnson, Proceedings of 15th ICPhS, 2385-2388, 2003.

[4] Boomershine *et al.*, In Avery *et al*. (Eds.), *Contrast in Phonology: Theory, Perception, Acquisition*, Mouton de Gruyter, 2008.

[5] Mielke, In Avery *et al.* (Eds.), *Contrast in Phonology: Theory, Perception, Acquisition*, Mouton de Gruyter, 2008.

[6] Whalen *et al.*, J. Phonetics, 25, 501-528, 1997.

[7] Celata, In Watkins *et al.*, *Recent Research in Second Language Phonetics/Phonology: Perception and Production*, Cambridge Scholars Publishing, 2009.

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

[9] Macmillan & Creelman, *Detection Theory: A User's Guide*, Psychology Press, 2004.