# The role of variability in non-native perceptual learning of a Japanese geminate-singleton fricative contrast

*Makiko Sadakata*[1] *& James M. McQueen*[1,2,3]

[1] Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University Nijmegen, The Netherlands
[2] Behavioural Science Institute, Radboud University Nijmegen, The Netherlands
[3] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

m.sadakata@donders.ru.nl, J.McQueen@pwo.ru.nl

## Abstract

The current study reports the enhancing effect of a high variability training procedure in the learning of a Japanese geminate-singleton fricative contrast. Dutch natives took part in a five-day training procedure in which they identified geminate and singleton variants of the Japanese fricative /s/. They heard either many repetitions of a limited set of words recorded by a single speaker (simple training) or fewer repetitions of a more variable set of words recorded by multiple speakers (variable training). Pre-post identification evaluations and a transfer test indicated clear benefits of the variable training.

**Index Terms**: Japanese fricatives, geminates, non-native speech learning, variability in training

## 1. Introduction

Learning to perceive a novel phonetic category as an adult is not very easy, yet it is one of the essential skills required when mastering a foreign language. Formation of perceptual categories is known to depend on the statistical information in training materials [1]. Furthermore, variability in training materials seems to play a crucial role in the formation of more robust perceptual categories. This has been demonstrated in perception of, e.g., English /r/ and /l/ by Japanese native speakers [2], Mandarin tones by English native speakers [3], and vowel length contrasts in Japanese by English speakers [4]. One of the possible underlying causes of this effect is an improved ability to generalize phonetic categories after exposure to a range of acoustic cues [2]. The current study investigates whether this effect extends to another situation, namely, Dutch natives' perception of a Japanese geminate-singleton fricative contrast.

Perceptual segmentation by Japanese natives does not always correspond to that by non-Japanese natives because Japanese natives segment speech into morae [5], whereas speakers of most other languages, including Dutch, do not [6]. This difference can lead to substantial segmentation differences, especially when a word contains special moraic features. One such case is the moraic obstruent /Q/ that occurs in geminated consonants [7]. For example, Japanese natives segment the word /sassu/ ("infer") into three morae (/sa.Q.su/) while non-Japanese natives would segment it into two syllables (/sa.su/). Japanese natives can thus easily distinguish /sassu/ from /sasu/ ("point") because /sasu/ has only two morae (/sa.su/). But this contrast can be problematic for non-Japanese natives because both words appear to have two syllables. Does variability during training help Dutch natives learn this kind of contrast?

Japanese geminates and singletons are not only perceptually but also acoustically distinct. Geminated consonants are often associated with longer preceding vowels and longer consonant durations than their singleton counterparts [8]. Previous studies indicated that Japanese listeners perceive this distinction in a categorical manner as a function of the duration of the critical consonant [9].

Gemination occurs with different Japanese consonants: stops (e.g., /p/, /k/), affricates (e.g., /ts/) and fricatives (e.g., /s/, /f/). A recent study has indicated that among these consonants, it is more difficult for non-Japanese natives to distinguish a fricative geminate /s/ followed by /u/ from its singleton counterpart (e.g., /assu/ vs. /asu/) than to distinguish other geminate-singleton contrasts involving stop consonants or consonants followed by /a/ [10]. Based on these findings, we decided to use the geminate-singleton contrast of /s/ followed by /u/ as learning material for Dutch native speakers. Dutch native speakers may well be sensitive to durational contrasts in speech signals because the Dutch vowel inventory includes long and short vowels [11]. However, Dutch does not contrast duration of consonants. Further, a recent study has indicated Dutch natives (non-musicians) perform relatively poorly in identifying a Japanese consonant duration contrast [12].

We compared the effects of two types of training material on perceptual learning: a simple training set which included more repetitions of a limited number of words recorded by a single speaker and a variable training set which included fewer repetitions of a more variable set of words recorded by multiple speakers. The effect of these two training methods was evaluated on identification accuracy and degree of transfer of learning. Three identification tests with different materials and procedures were performed to this end: training, identification tests (pre- and post-training tests) and a transfer identification test.

## 2. Methods

### 2.1. Participants

Thirty native Dutch speakers recruited from the participant pool of the Max Planck Institute for Psycholinguistics took part. They were randomly assigned to two groups: simple and variable. None of them had had any intensive exposure to Japanese.

### 2.2. Stimuli

Table 1 provides a list of naturally spoken Japanese words used for the three tests. The pre/post-training identification test and the transfer identification test used minimal trios contrasting a singleton (CVCV), a geminate (CVCCV), and a singleton with preceding long vowel (CV:CV), while the training used minimal pairs contrasting the singleton (CVCV) and the geminate (CVCCV). The pitch relationship between the first and the second CV was fixed to high-low. All recordings were first low-pass filtered at 5000 Hz; average sound levels were normalized to 70dB using Praat [13].

### 2.2.1. *Identification test*

The stimuli used for the pre- and post-training identification tests consisted of 11 minimal trios, which contrasted three word types (singleton, geminate, long vowel) with the fricative /s/. One trio did not include an initial C (i.e., VCV, VCCV, V:CV) and served as example categories during the learning task. Altering the first C created the other 10 trios (CVCV-CVCCV-CV:CV). A female Japanese native (F1) recorded the material.

### 2.2.2. *Identification training*

The set of stimuli used for the simple training condition was identical to that used for the identification test but without the long vowel condition. The pair without the first C spoken by F1 was used as example categories during the learning task for this condition. The following training task used two minimal pairs with a CVCV structure per day, which summed to 10 minimal pairs after five training sessions. We refer to this set as SF1 (simple F1). For the variable training condition, the pair without the first C (/asu/-/assu/) recorded by five speakers (3 females, 2 males) was presented as example categories during the learning task. For the training sessions, we created 40 word pairs with a CVCV structure by altering the first C and the last V. Five speakers recorded these pairs, which resulted in 200 pairs in total. We refer to these speakers as VF1, VF2, VF3, VM1, VM2, respectively. Forty minimal pairs (8 training pairs recorded by different speakers per day) were used per participant. We used three orders to present speakers (VF1-VM1-VF2-VM2-VF3, VF2-VM1-VF3-VM2-VF1, VF3-VM2-VF1-VM1-VF2).

### 2.2.3. *Transfer identification test*

The stimuli of the transfer test included three new types of minimal trios, based on the stop /k/ (/paku/-/pakku/-/pa:ku/), the affricate /ts/ (/patsu/-/pattsu/-/pa:tsu/), and the vowel /e/ (/pase/-/passe/-/pa:se/). There were also familiar trios, either embedded in a sentence (/kore ha pasu desu/ *this is pasu*), or spoken by a new male speaker (M3).

## 2.3. Procedure

### 2.3.1. *Identification and transfer identification test*

The identification and transfer tests both consisted of a learning task followed by an identification task. During the learning task, participants were presented with six repetitions of three example categories, e.g., /asu/-/assu/-/a:su/. Each sound was presented along with visual number *1*, *2* and *3*, each associated with a labeled key on the keyboard with an inter-stimulus interval (ISI) of 2000 ms. We presented visual numbers and sound categories in three combinations. The combination was kept constant for each participant across the three tests in each of the five experimental sessions. For example, some participants learned to associate category /asu/ with *1* while others learned to associate it with *2* or *3*, etc. During the identification task, one of the CVCV, CVCCV and CV:CV words was presented per trial and participants pressed the key 1, 2 or 3 to indicate their categorical judgments. The inter-trial interval (ITI) was set to 1500 ms after the response key-press. The identification test had 90 trials and the transfer test had 450 trials.

Table 1. *List of words used for the three tests.*

| Test | Condition | Example | First C | Last V |
|---|---|---|---|---|
| **Identifica-tion** | | asu - assu - a:su | b, d, g, k, m, n, p, t, w, y | u |
| **Training** | Simple | asu - assu | | u |
| | Variable | asu - assu | | a, i, u, o |
| **Transfer** | New $c_2$ | aku - akku - a:ku | b, d, g, s, m, n, p, t, w, y | u |
| | New $c_2$ | atsu - attsu - a:tsu | b, d, g, s, m, n, p, k, w, y | u |
| | New $v_2$ | esu - essu - e:su | b, d, g, k, m, n, p, t, w, y | e |
| | New speaker | asu - assu - a:su | b, d, g, k, m, n, p, t, w, y | u |
| | Long sentence | asu - assu - a:su | | u |

### 2.3.2. *Training*

Each training session consisted of a learning task followed by an identification task with visual feedback on the correctness of responses. During the learning task, participants were presented with two example categories, /asu/-/assu/. The simple condition used the same example pair by F1 for all the five training sessions while the variable condition used example pairs recorded by five different speakers (one speaker per session). The presentation of example minimal pairs was repeated six times. Each category was presented along with visual number *1* or *2* with an ISI of 2000 ms. During the identification task, a CVCV word was presented per trial and participants pressed the key *1* or *2* to indicate their categorical judgments. Feedback on the correctness of response was provided as visual letters for 2000 ms after the response press. The ITI was set to 1500 ms. Each training session consisted of 5 blocks of 32 trials each.

# 3. Results

Responses with a reaction time longer than 3 standard deviations from the mean were identified as outliers. 2.1 % of the responses were thus discarded from the analyses.

Figure 1 shows the identification accuracy of the pre-test (before the first training) as well as of the five post-tests for the simple and variable groups. A t-test confirmed no significant difference between the groups' correct response rates for the first pre-test ($t(28) = -1.59$, $p = 0.124$), indicating that the task performance of two groups was equivalent before the training. First, training effects on absolute correct response rate were examined using a two-way ANOVA with Training session (continuous variable) and Group (simple/variable) as independent variables and participants as random variable. There was a significant effect of Training session ($F(1, 118)=79.7$, $p<.0001$) but no significant effect of Group ($F(1,28)=1.37$, $p=.25$) and no interaction ($F(1,118)=0.13$, $p=.72$). The relative improvement rate was then calculated by subtracting the correct response rate of the first pre-test from that of five post-tests. An analogous two-way ANOVA indicated main effects of Training session (F(1, 118)=79.7, p<.0001) and Group (F(1,28)=10.6, p=.0030), without a significant interaction (F(1,118)=0.13, p=.72). This latter analysis indicates that both groups increased their response accuracy over the course of the five training sessions and that the participants in the variable condition showed significantly higher identification accuracy for all post-tests. However, even after the five training sessions Dutch participants did not reach the 85% accuracy level, which is well below native Japanese identification accuracy.

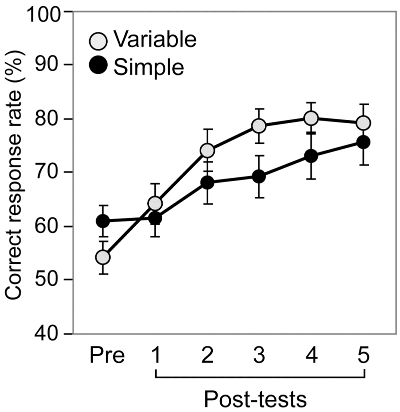


Figure 1: *Correct response rate for pre- and post-tests. Error bars indicate standard errors.*

### 3.1.1. Training

Responses with a reaction time longer than 3 standard deviations from the mean were identified as outliers. 1.9 % of the responses were thus discarded from the analyses.

Figure 2a shows the correct response rate for the first training sessions (treated as a continuous variable) for the four types of training materials. A two-way ANOVA with Block (continuous variable) and Training material (Simple F1, Variable F1, Variable F2, Variable F3) as independent variables and participants as random variables revealed a significant main effect of Training material (F(3,26)=9.0, p=.0003) with no significant main effect of Block (F(1,116)=0.8, p=.038) and no interaction (F(3,26)=1.3, p=.28). A multiple comparison confirmed that the correct response rate of SF1 was significantly lower than that of VF2 and VF3 (p<.05). The correct response rate of VF1 did not significantly differ from any of the other three conditions.

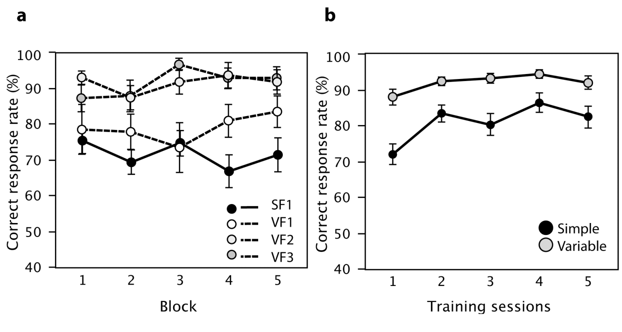


Figure 2: *a) Correct response rate for 5 blocks of the first training session and b) correct response rate over the five sessions. Error bars indicate standard errors.*

Figure 2b shows the correct response rate as a function of the five training sessions for two groups (simple and variable, respectively). A two-way ANOVA with Group (simple/variable) and Training Session (a continuous variable) as independent variables and participants as random variables indicated main effects of Group (F(1,28)=24.3, p< .0001) and Training Session (F(1,118)=14.5, p=.0002) without a significant interaction (F(1,118)=2.6, p=.11). The main effect of Training Session indicates that both groups improved their learning over the course of five training sessions. More importantly, the main effect of Group with significantly higher correct response rate for the variable group indicates that, despite the wider variability included in training materials, the variable group performed the task better.

### 3.1.2. Transfer test

Due to a technical error, the data of one participant from the variable group was lost for the transfer test and excluded from further analyses. A repeated measure ANOVA with Group (simple/variable) as between-subjects independent variable and Condition (5 conditions) as within-subjects independent variable indicated significant main effects of Group (F(1,27) =5.3, p=.029) and Condition (F(4,24)=18.5, p<.0001) without a significant interaction. As can be seen in Figure 3, the variable group responded more accurately to all transfer conditions. Post-hoc comparisons across the five conditions revealed that response accuracy in the sentence and new speaker conditions was significantly higher than in the other three conditions (p<.05).

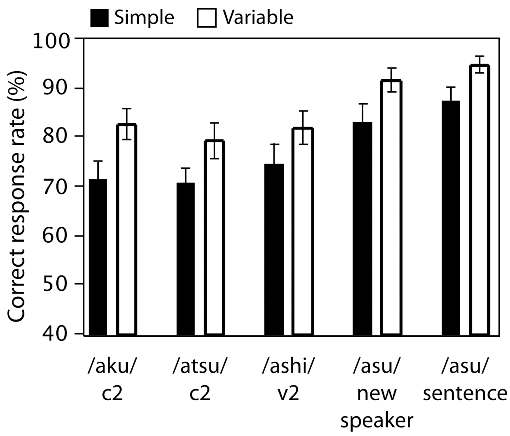


Figure 3: *Correct response rate in the five transfer conditions.*

## 4. Discussion

In the current study, the effects of variability of training materials on auditory category learning were investigated. The simple group heard the stimuli of the identification test more often than the variable group during training while the variable group was exposed to a wider range of example words produced by multiple speakers. The variable group outperformed the simple group with regard to their response accuracy in all the post-tests as well as in the transfer test. This indicates that the high-variability training method [2-4,14] is also effective for learning of the Japanese geminate-singleton fricative contrast /s/ by Dutch natives. The benefit of the high-variability training method was already there in the first post-test results, indicating that the bulk of this effect took place during the first training session.

A crucial acoustic cue that distinguishes geminate-singleton consonants is duration, although other acoustical features such as loudness change also contribute to the distinction [15]. For example, the acoustic differences between geminate-singleton consonants are mainly manifested in length of the preceding vowel and the critical consonant [8], various durational ratios such as closure to word duration [16] and closure to preceding mora duration or to following vowel [17]. It is known that learning of such Japanese durational contrasts is challenging for non-Japanese natives whose mother tongues do not require phonemic recognition based on duration, such as English and German [10,14,18]. However, recent studies with English and German natives reported that training improves perception of Japanese durational contrasts to some extent [4,14,18]. Unlike English, Dutch has phonemic vowel contrasts based (at least in part) on duration [11]. However, perceptual sensitivity to vowel duration seemed not to result in an immediate advantage when perceiving Japanese geminate-singleton consonants because response accuracy in the pre-test of our study was rather poor. This is in line with the previous finding that demonstrated limited generalization about a durational contrast by non-Japanese natives: English natives who were trained to perceive a Japanese vowel durational contrast did not significantly improve their perception of other Japanese durational contrasts, such as the geminate-singleton contrast [14]. If there is any advantage for Dutch natives, it may be in the learning rate and/or the transfer phase (but this cannot be detected in the current experiment).

The variable group in this study performed fairly well on the transfer test for all conditions. It is perhaps unsurprising that the variable group generalized their training to a new successive vowel (/e/) and to a new speaker (M3) because the variable materials varied these factors in a similar manner. What is more interesting is the relatively high performance accuracy on the two new consonants (/k/, /ts/). Because the critical consonant /s/ was kept constant during the experiment, being able to perform the task well on these new consonant conditions requires true generalization of learned knowledge, that is, knowledge that a consonantal distinction can be based on durational cues. It is also interesting that response accuracy was rather good for both groups in the sentence condition. The carrier sentence might have provided richer context information that could be used to improve task performance. Indeed, embedding target words in a sentence is reported to be more effective when learning durational categories [19,20,21].

Speaker F1 appeared to be the most difficult speaker among the six speakers (including M3). Figure 2a indicates that participants had a hard time performing the identification task in the first training session even with feedback. However, a control study confirmed that Japanese natives identified the geminate-singleton-long vowel contrast spoken by F1 fairly well without any training (N=10, 91%), confirming the validity of this material. Interestingly, variability in material might have been helpful here because accuracy for the VF1 condition did not differ significantly from that for the VF2 and VF3 conditions.

# 5. Conclusion

The current study demonstrated that perception of the Japanese geminate-singleton /s/ contrast is challenging for Dutch natives but that it is possible for them to improve with training. Furthermore, the pre- and post-training identification evaluations and the transfer test indicated clear benefits of the variable training. Variability during training thus seemed to help Dutch natives learn this non-native durational contrast.

# 6. Acknowledgements

The authors are grateful for support provided by Leonhaldt Rau. We are also grateful for useful comments from Mirjam Broersma and Holger Mitterer.

# 7. References


[1] J. Maye, J.F. Werker, and L. Gerken, "Infant sensitivity to distributional information can affect phonetic discrimination", Cognition, 82(3): B101-B111, 2002.

[2] J.S. Logan, Liverly, S. E., Pisoni, D. B., "Training Japanese listeners to identify English /r/ and /l/: A first report," Journal of Acoustical Society of America, 89(2):874-886, 1991.

[3] Y. Wang, M.M. Spence, A. Jongman, and J.A. Sereno, "Training American listeners to perceive Mandarin tones," Journal of the Acoustical Society of America, 106(6):3649-3658, 1999.

[4] Y. Hirata, E. Whitehurst, and E. Cullings, "Training native English speakers to identify Japanese vowel length contrast with sentences at variet speaking rates," Journal of Acoustical Society of America, 121(6):3837-3845, 2007.

[5] T. Otake, G. Hatano, A. Cutler, and J. Mehler, "Mora or Syllable - Speech Segmentation in Japanese," Journal of Memory and Language, 32 (2), pp. 258-278, 1993.

[6] J. Vroomen, M. van Zon, and B. de Belder, "Cues to speech segmentation: evidence from juncture misperceptions and word spotting," Memory & Cognition, 24, 1996, pp. 744-755.

[7] T.J. Vance, "An introduction to Japanese phonology," State University of New York Press, 1987.

[8] J. Kingston, S. Kawahara, D. Chambless, D. Mash, and E. Brenner-Alsop, "Contextual effects on the perception of duration," Journal of Phonetics, 37(3):297-320, 2009.

[9] M. Inouchi, M. Kubota, K. Ohta, Y. Shirahama, A. Takashima, T. Horiguchi, and E. Matsushima, "Human auditory evoked mismatch field amplitudes vary as a function of vowel duration in healthy first-language speakers," Neuroscience Letters, 366(3):342-346, 2004.

[10] D.M. Hardison, and M.M. Saigo, "Development of perception of second language Japanese geminates: Role of duration, sonority, and segmentation strategy," Applied Psycholinguistics, 31(1):81-99, 2010.

[11] C. Gussenhoven, "Dutch," Journal of the International Phonetic Association, 22(1-2):45-47, 1992.

[12] M. Sadakata, and K. Sekiyama, "Enhanced perception of various linguistic features by musicians: A cross-linguistic study," Acta Psychologica, in press.

[13] P. Boersma, and Weenink, D., "Praat, a system for doing phonetics by computer" Glot International, 5(9-10):341-345, 2001.

[14] K. Tajima, H. Kato, A. Rothwell, R. Akahane-Yamada, and K.G. Munhall, "Training English listeners to perceive phonemic length contrasts in Japanese," Journal of Acoustical Society of America, 123(1):397-413, 2010.

[15] K. Idemaru, and S. G. Guion, "Acoustic covariants of length contrast in Japanese stops," Journal of the International Phonetic Association, 38:167-186, 2008.

[16] Y. Hirata, "Effects os speaking rate on the single/geminate stop distinction in Japanese," Journal of Acoustical Society of America, 118(3):1647-1660, 2005.

[17] K. Idemaru, and S. Guion-Anderson, "Relational timing in production and perception of Japanese singleton and geminate stops," Phonetica, 67:25-46, 2010.

[18] H. Menning, S. Imaizumi, P. Zwitserlood, and C. Pantev, "Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language," Learning & Memory, 9(5):253-267, 2002.

[19] M. Hisagi, and W. Strange, "Perception of Japanese temporally-cued contrasts by American English listeners," Language and Speech, 54(2):241-264, 2011.

[20] M. Sonu, K. Tajima, H. Kato, and Y. Sagisaka, "The effect of a word embedded in a sentence and speaking rate variation of the perceptual training of geminate and singleton consonant distinction," Proceedings of INTERSPEECH-2010, 1850-1853, 2010.

[21] Y. Hirata, "Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts," Journal of the Acoustical Society of America, 116(4):2384-2394, 2004.