High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates

Makiko Sadakata[a)]
Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University Nijmegen, P.O. Box 9104 6500 HE, Nijmegen, The Netherlands

James M. McQueen[b)]
Donders Institut for Brain, Cognition and Behaviour, Centre for Cognition and Behavioural Science Institute, Radboud University Nijmegen, P.O. box 9104, 6500 HE, Nijmegen, The Netherlands

(Received 11 May 2012; revised 4 June 2013; accepted 12 June 2013)

This study reports effects of a high-variability training procedure on nonnative learning of a Japanese geminate-singleton fricative contrast. Thirty native speakers of Dutch took part in a 5-day training procedure in which they identified geminate and singleton variants of the Japanese fricative /s/. Participants were trained with either many repetitions of a limited set of words recorded by a single speaker (low-variability training) or with fewer repetitions of a more variable set of words recorded by multiple speakers (high-variability training). Both types of training enhanced identification of speech but not of nonspeech materials, indicating that learning was domain specific. High-variability training led to superior performance in identification but not in discrimination tests, and supported better generalization of learning as shown by transfer from the trained fricatives to the identification of untrained stops and affricates. Variability thus helps nonnative listeners to form abstract categories rather than to enhance early acoustic analysis.
 [http://dx.doi.org/10.1121/1.4812767]

PACS number(s): 43.71.Ft, 43.71.Bp, 43.71.Es, 43.71.Hw [BRM] Pages: 1324–1335

I. INTRODUCTION

Learning to perceive novel phonetic categories is one of the essential skills we need to acquire when mastering a foreign language. There is a growing body of evidence that exposure to variable nonnative speech materials results in more successful perceptual learning. For example, Logan *et al*. (1991) demonstrated that native speakers of Japanese who were trained to make the English /r/-/l/ distinction with materials with multiple voices and variable word forms showed better learning in an identification test than those in the study by Strange and Dittmann (1984) who were trained to discriminate the same contrast with limited variability. This effect has been replicated with different phonetic contrasts, such as the Mandarin tone contrast by native speakers of English (Wang *et al*., 1999) and vowel length contrasts in Japanese by native speakers of English (Hirata *et al*., 2007).

One reason why variability in training materials may strengthen the process of category formation is that it may help learners develop abstract representations that can accommodate a wider range of examples (e.g., Logan *et al*., 1991). Alternatively, however, experiencing variability might enhance one's sensitivity to novel types of speech signals at a pre-categorical level, which in turn could contribute to better identification. The present study compared these two hypotheses, primarily by measuring the effects of high- and low-variability training on two perceptual tasks. The results of an identification task which required listeners to use categorical information (2 alternative-forced choice, 2AFC, e.g., Pisoni 1977; Bradlow *et al*., 1997) were compared with those of a discrimination task which can be performed without recourse to categorical information (4-interval 2-alternative forced choice, 4I2AFC, e.g., Gerrits and Schouten, 2004). If high-variability training enhances only the formation of phonetic categories, it may lead to improved identification but not discrimination. In contrast, if high-variability training enhances pre-categorical sensitivity that then in turn supports formation of phonetic categories, there should be increases in accuracy arising from high-variability training in both tasks. Through contrasting these two hypotheses, therefore, we attempted to identify what underlies the benefit that high-variability training has on nonnative speech perception.

We also examined individual differences in learning nonnative speech sounds. Sensitivity to acoustical signals varies considerably among individuals. For example, musicians are known to be more accurate at encoding pitch and timing patterns of speech sounds at a very early stage of perception and this is reflected when they perceive different speech materials (Besson *et al*., 2007; Wong *et al*., 2007; Sadakata and Sekiyama, 2011). Using various phonetic contrasts, however, Sadakata and Sekiyama (2011) reported that musicians who were better at discriminating vowel contrasts than nonmusicians were not necessarily better at identifying

[a)]Author to whom correspondence should be addressed. Electronic mail: m.sadakata@donders.ru.nl
[b)]Also at Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands and Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University Nijmegen, The Netherlands.

 0001-4966/2013/134(2)/1324/12/$30.00

the same phonetic categories. This means that good sensitivity to acoustic features alone does not necessarily lead to good identification. Nevertheless, better discrimination ability, which provides the listener with more detailed acoustic information about the target sound, could very well be useful when he or she is learning to extract and generalize essential characteristics of a new phonetic category. In the present study, therefore, we asked not only what the source of the benefit arising from high-variability training is, but also whether there are individual differences among listeners in discrimination ability and whether those differences predict listener success in the training procedure.

We trained native speakers of Dutch to identify a nonnative durational contrast in spoken Japanese with two levels of variability, and tested the effects of this training on identification (2AFC) and discrimination (4I2AFC) accuracy. More specifically, the contrast concerned the distinction between Japanese singleton and geminate fricative consonants, such as /ss/ and /s/. Although our hypotheses could be tested using a variety of speech categories, this contrast was chosen because it is less studied in comparison, for example, to tone contrasts, and because it is based on durational rather than pitch or spectral differences. Speakers of all languages should have been trained to deal with speech timing information to some extent, simply because speech evolves over time. Nevertheless, learning of the Japanese geminate contrast has been reported to be challenging for nonnative speakers of Japanese, such as English and German speakers (Menning *et al.*, 2002; Hardison and Saigo, 2010; Tajima *et al.*, 2010). It has also been shown, however, that training improves perception of this contrast to some extent (Hirata *et al.*, 2007; Menning *et al.*, 2002). The question we asked, therefore, was whether larger improvements in learning about Japanese geminates would be found with higher variability training.

Gemination occurs with various consonants in Japanese: stops (e.g., /p/, /k/), affricates (e.g., /ts/), and fricatives (e.g., /s/, /f/). People who are not native speakers of Japanese find the fricative geminate /s/ followed by /u/ more difficult to distinguish from its singleton counterpart (e.g., /assu/ vs /asu/) than other geminate-singleton contrasts involving stop consonants or consonants followed by /a/ (e.g., /atta/ vs /ata/; Hardisson and Saigo, 2010). Based on this finding, we decided to use the geminate-singleton contrast of /s/ followed by /u/ as the training material.

Among several acoustic differences between Japanese geminates and singletons (Han, 1992; Idemaru and Guion, 2008), local timing features appear to be the most important. Primary cues include the ratio of the closure to word duration in the case of stops (Hirata and Whiton, 2005; Amano and Hirata, 2010), and the relative duration of the preceding vowel and the critical consonant (Kingston *et al.*, 2009). Kingston *et al.* (2009) showed that identification of nonnative (Japanese) geminate stop consonants (in a 2AFC task) was modulated by the way vowel and consonant durations covary in the native language of different groups of listeners (Italian, English, and Norwegian). Importantly, however, Kingston *et al.* observed no differences as a function of native language when their participants identified nonspeech analogs of the geminate stimuli and when participants performed discrimination tasks (2I2AFC) on speech and on nonspeech materials. These findings support the view that identification and discrimination tasks can tap into different stages in perception (cf. Gerrits and Schouten, 2004; Sjerps *et al.*, 2013), and suggest that linguistic knowledge is applied only at a later stage in perception than that tapped into by discrimination tasks.

Knowledge about duration covariance acquired in the long-term (through native-language exposure) thus appears not to influence perception of nonspeech stimuli. The question that then arises is whether short-term geminate training influences perception of nonspeech analogs of geminates. In the present study we therefore used speech materials and nonspeech analogs of those materials in the identification and discrimination tasks. If training influences low-level domain-general auditory processes (i.e., those responsible for the processing of relative duration information that is not specific to speech), then there should be effects of training on both speech and nonspeech test materials, with potentially also a difference between the high- and low-variability training conditions. If, however, training has an effect at a higher, domain-specific level of processing, there should be no transfer of training effects to nonspeech materials.

To summarize, we studied the effect of variability during perceptual training on identification and discrimination of speech and nonspeech materials. The effects of two types of Japanese geminate training material were compared across two participant groups. One group received a low-variability training set which included more repetitions of a limited number of words recorded by a single speaker. The other group received a high-variability 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 by measuring improvements in identification accuracy (on both trained and new natural speech materials), in discrimination accuracy on stimuli from synthesized continua (based on speech and on nonspeech analogs), and by testing for changes in the sharpness of the categorization functions of the synthesized speech and nonspeech continua. Table I shows the structure of the five experimental sessions. Different sets of tests were administered each day in order to address three research questions.

Our primary question was whether variability enhances formation of abstract categorical representations of geminates. If so, participants who are trained using high-variability geminates should show better identification accuracy than participants trained with low variability. If high-variability participants do not also show better geminate discrimination accuracy, as measured by means of a discrimination task which focuses on pre-categorical processing (4I2AFC; Gerrits and Schouten, 2004; Sjerps *et al.*, 2013), then the high-variability enhancement is not likely to be due to changes in perceptual sensitivity. In contrast, if high-variability participants were to show better identification and discrimination accuracy, then this would indicate that the benefit of high variability is due, at least in part, to enhanced pre-categorical sensitivity.

Furthermore, if high-variability participants have formed an abstract geminate category they should be able to

TABLE I. The structure of the five experimental sessions.

Day 1	Day 2	Day 3	Day 4	Day 5
Identification	Identification	Identification	Identification	Identification
Discrimination (s)	*Training*	*Training*	*Training*	*Training*
Discrimination (n)	Identification	Discrimination (s)	Identification	Transfer
Categorization (s)		Discrimination (n)		Discrimination (s)
Categorization (n)		Identification		Discrimination (n)
Training				Categorization (s)
Identification				Categorization (n)
				Identification
75 min	30 min	50 min	30 min	90 min

Note: s = speech materials, n = nonspeech materials.

transfer what they have learned about fricative durations to other stimuli. Testing identification accuracy on new materials indicates how general the learning effect is (e.g., Bradlow and Pisoni, 1999; Tajima *et al.*, 2010). If the participants in the high-variability group have more successfully abstracted knowledge about geminates, as opposed to having learned specific fricative timing information more successfully, they should perform better than the low-variability group on untrained materials. We therefore tested for transfer of learning to new consonants, new vowels, a new speaker, and a new context.

The second question examined individual differences. Does a participant who shows superior performance in discrimination show better improvement in identification? This question was examined by asking whether pre-test discrimination accuracy results predicted the amount of improvement (identification accuracy over the course of the 5 days of training).

The third research question concerned whether perceptual learning of speech material generalizes to nonspeech materials. To answer this question, results for speech and nonspeech materials were compared in the discrimination and categorization tasks.

II. METHOD

A. Participants

Thirty native speakers of Dutch recruited from the participant pool of the Max Planck Institute for Psycholinguistics took part (18 females and 12 males, average age of 21.3 yr old). They were randomly assigned to two groups: low-variability and high-variability. Although the majority of the participants spoke multiple languages at different fluency levels, none of them had had any substantial exposure to Japanese. All participants indicated their self-evaluated fluency level on a scale from 1 (not fluent at all) to 5 (very fluent) with regard to each of their L2s. The reported L2s included English (N = 30, fluency level 3.5 – 5), German (N = 23, fluency level 1 – 3), French (N = 22, fluency level 1 – 3), Spanish (N = 5, fluency level 1 – 4) and Hungarian (N = 1, fluency level 1). Participants received 60 euros after taking part in five training sessions. There was no dropout.

B. Stimuli

1. Natural speech

The identification test, the training procedure and the transfer test (see Table I) used naturally spoken materials. Table II provides a list of the words used in all three of these tests. The pre- and post-training identification test and the transfer identification test used minimal trios contrasting the Japanese singleton (CVCV), geminate (CVCCV), and singleton with preceding long vowel (CV:CV), while the training used minimal pairs contrasting the Japanese singleton (CVCV) and geminate (CVCCV). The preceding long vowel condition, which was used in previous studies (e.g., Hardison *et al.*, 2010), was included in order to make the task more challenging. The pitch-accent relationship between the first and the second CV was fixed to high-low. These materials were spoken by six native speakers of Japanese (3 females and 3 males). All recordings were first low-pass filtered at 5000 Hz and average sound levels were normalized to 70 dB using Praat (Boersma and Weenink, 2008).

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). One of the female speakers (F1) recorded these materials.

The set of stimuli used for the low-variability training condition were identical to that used for the identification test but without the long vowel condition. The pair without the first C spoken by speaker 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 LF1 (low-variability, speaker F1). For the high-variability training condition, the pairs without the first C (/asu/-/assu/) recorded by F1 and by four of the other speakers (2 females, 2 males) were 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. All five of these speakers recorded these pairs, which resulted in

TABLE II. Summary of stimuli in the identification test, training, and the transfer test.

Test	Condition	Example	c_1	v_2	Speakers
Identification trio		asu - assu - a:su	b, d, g, k, m, n, p, t, w, y	u	F1
Training pair	low-variability	asu - assu	b, d, g, k, m, n, p, t, w, y	u	F1
	high- variability	asu - assu	b, d, g, k, m, n, p, t, w, y	a, i, u, o	F1, F2, F3, M1, M2
Transfer trio	new c_2 (stop)	aku - akku - a:ku	b, d, g, s, m, n, p, t, w, y	u	F1
	new c_2 (affricate)	atsu - attsu - a:tsu	b, d, g, s, m, n, p, t, w, y	u	F1
	new v_2	esu - essu - e:su	b, d, g, k, m, n, p, t, w, y	e	F1
	long sentence	asu - assu - a:su	b, d, g, k, m, n, p, t, w, y	u	F1
	new speaker	asu - assu - a:su	b, d, g, k, m, n, p, t, w, y	u	M1

200 pairs in total. We refer to these as HF1, HF2, HF3, HM1, HM2, respectively. Among them, 40 minimal pairs (8 training pairs recorded by different speakers per day) were used per participant. We used three orders to present speakers (HF1-HM1-HF2-HM2-HF3, HF2-HM1-HF3-HM2-HF1, HF3-HM2-HF1-HM1-HF2). Example presentation patterns are given in Table III. Average durations (ms) of each component used for the low- and high-variability training materials and for the transfer test materials are given in Table IV.

The stimuli of the transfer test included five new types of minimal trios: (1) stop /k/ (/paku/-/pakku/-/pa:ku/), (2) affricate /ts/ (/patsu/-/pattsu/-/pa:tsu/), (3) vowel /e/ (/pase/-/passe/-/pa:se/), and familiar trios either (4) embedded in a sentence (/kore ha *pasu* desu/ *this is pasu*) or (5) spoken by the third male speaker (M3).

2. Synthesized speech

Categorization and discrimination of synthesized continua of speech and nonspeech sounds were also tested (see Table I). Table V summarizes the durations of the intervals in the synthesized stimuli.

Construction of synthesized stimuli was similar (but not identical) to that in Kingston *et al.*, (2009). For the speech materials, /asu/ - /assu/ continua were created by changing the duration of the preceding vowel /a/ (V_1), the critical consonant /s/ (C) and the final vowel /u/ (V_2). The original sounds of /a/, /s/ and /u/ were taken from the natural utterance of /asu/ spoken by F1 with durations of 63, 105, and 63 ms, respectively. The duration of /s/ was modulated from 60 to 150 ms in seven equal steps of 15 ms and V_1 duration was scaled to 43, 63, and 89 ms, by cutting out periods of voicing using Praat (Boersma and Weenink, 2008). All combinations of C and V durations were created, which resulted in a set of 21 stimuli. Onsets and offsets of each element were ramped (5 ms) to ensure that they were zero crossings at the points of concatenation. The 21 stimuli were grouped into short-vowel, medium-vowel, and long-vowel continua.

The nonspeech continuum consisted of three elements that were analogs to the V1CV2 structure of the synthesized speech material. A filtered square wave was used for the consonant portion and an anharmonic complex of sine waves was used to represent the two vocalic portions. The F0 of the filtered square wave was 100 Hz and the first 50 odd harmonics were included with amplitudes of 1/harmonic number (i.e., 1/1, 1/3, 1/5, etc.). The anharmonic complex was made of 50 sine waves with frequencies ranging from 100 to 16 000 Hz. The frequencies were separated by equal natural log intervals (0.101503) and added with amplitude ratios of 1/(2*component number + 1) (i.e., 1/3, 1/9, 1/19, etc.). The duration of each portion was manipulated in the same manner as in the speech materials. For more complete description of the stimuli, please refer to Kingston *et al.* (2009).

The categorization and discrimination tests used two continua, one based on consonants and one on vowels, for both the speech and the nonspeech materials. The durations of the consonant continua (speech and nonspeech analog) were identical to the identification stimuli in Kingston *et al.* (2009), with a range of 60–150 ms (step size = 15 ms) and with fixed V duration of 70.5 ms. The vowel and vowel-analog continua included various V_1 durations with a range of 48–93 ms in 15 ms steps followed by a fixed C with a duration of 97.5 ms.

C. Procedure

Experiment sessions took place over 5 days with a maximum duration of 2 days between sessions. Total duration of the experiment ranged from 5 to 9 days. As shown in

TABLE III. Example of stimulus presentation patterns for the low- and high-variability groups.

		Day 1	Day 2	Day 3	Day 4	Day 5
Low-variability	speaker	F1	F1	F1	F1	F1
	stimuli	pasu, kasu	basu, tasu	nasu, hasu	masu, dasu	gasu, yasu
High-variability	speaker	F1	M1	F2	M2	F3
	stimuli	pasa, kasa, pasu, kasu, pase, kase, paso, kaso	basa, tasa, basu, tasu, base, tase, baso, taso	nasa, hasa, nasu, hasu, nase, hase, naso, haso	masa, dasa, masu, dasu, mase, dase, maso, daso	gasa, yasa, gasu, yasu, gase, yase, gasa, yasa

TABLE IV. Average duration of each component used in the low- and high-variability training and transfer test materials (ms, numbers in brackets are standard deviations).

Condition	Speaker	Type	Total	C1	V1	C2	V2
Low- variability	F1	G	444.0	35.2	66.6	212.6	132.6
			(17.9)	(18.2)	(6.6)	(9.3)	(24.4)
		S	420.5	44.4	81.7	149.8	148.5
			(25.0)	(16.9)	(9.3)	(11.7)	(25.8)
		L	432.0	35.8	99.6	142.6	158.0
			(44.5)	(12.1)	(12.4)	(14.5)	(41.7)
High- variability	F1	G	462.3 (40.8)	41.0 (23.7)	73.8 (10.4)	229.1 (18.3)	119.3 (23.6)
		S	412.0 (46.8)	46.0 (24.5)	87.6 (14.2)	136.2 (14.7)	143.3 (23.4)
		L	483.7 (44.9)	43.8 (20.7)	136.7 (14.4)	164.9 (9.6)	114.2 (26.8)
	F2	G	495.3 (37.1)	49.1 (16.5)	82.5 (9.4)	276.3 (30.8)	84.5 (15.1)
		S	340.4 (28.7)	45.1 (18.0)	74.1 (11.8)	139.3 (19.6)	77.9 (14.6)
	F3	G	529.2 (44.4)	44.6 (30.8)	83.9 (14.0)	270.1 (15.8)	95.6 (16.4)
		S	352.2 (40.7)	40.4 (23.2)	63.2 (14.8)	132.4 (14.8)	85.9 (12.6
	M1	G	556.3 (40.9)	45.2 (19.3)	105.6 (12.0)	241.9 (22.8)	164.6 (34.3)
		S	428.2 (49.6)	40.1 (14.6)	90.1 (13.9)	120.4 (17.1)	176.7 (38.3)
	M2	G	533.9 (53.4)	58.0 (24.2)	77.9 (13.8)	263.9 (29.1)	101.6 (18.9)
		S	387.0 (40.0)	56.3 (18.6)	63.3 (15.6)	143.3 (25.5)	100.3 (20.9)
Transfer	F1 /aku/	G	453.1	45.1	77.6	219.6	137.2
			(14.1)	(16.8)	(14.1)	(17.9)	(14.7)
		S	445.5	58.1	93.8	141.9	151.8
			(17.5)	(15.2)	(14.1)	(17.9)	(14.6)
		L	481.9	64.4	171.8	139.7	130.6
			(54.7)	(19.0)	(16.4)	(18.6)	(12.7)
	F1 /atsu/	G	418.1	56.1	85.3	241.2	74.7
			(53.8)	(29.9)	(12.0)	(16.2)	(17.1)
		S	389.9	41.8	94.3	144	99.6
			(35.6)	(24.2)	(12.6)	(42.4)	(19.0)
		L	417.6	43.4	169.2	151.4	74.9
			(47.4)	(14.5)	(10.9)	(41.4)	(26.0)
	F1 /ashi/	G	388.7	31.6	81.4	212.2	98.9
			(44.2)	(22.5)	(9.4)	(13.9)	(16.8)
		S	376.4	37.3	85.6	141.8	108.5
			(28.4)	(18.7)	(12.4)	(20.6)	(13.4)
		L	450.4	44.7	144.4	143.8	117.4
			(23.3)	(18.7)	(15.9)	(11.5)	(17.5)
	F1 Sentence	G	424.0	26.0	78.6	234.7	84.7
			(14.0)	(8.7)	(14.7)	(15.9)	(11.8)
		S	309.7	36.8	71.8	126.3	74.8
			(25.1)	(18.1)	(13.3)	(11.3)	(6.8)
		L	408.6	35.7	165.9	126.2	80.8
			(15.1)	(13.5)	(18.6)	(11.1)	(13.1)
	M3	G	493.5	45.2	85.9	252.9	116.6
			(43.3)	(29.3)	(10.5)	(15.5)	(20.6)
		S	359.7	45.6	68.2	142.4	107.1
			(40.12)	(22.7)	(12.1)	(10.1)	(23.9)
		L	487.8	45.1	173.6	165.8	103.4
			(42.1)	(19.5)	(15.2)	(18.0)	(31.5)

Note: G = geminate, S = singleton, L = long vowel.

TABLE V. Durations of the components in the synthesized stimuli (ms).

Test	Condition	/a/	/s/	/u/
Categorization	short V_1, medium V_1, long V_1	49, 63, 89	60, 75, 90, 105, 120, 135, 150	105
Discrimination	consonant contrast	70.5	60, 75, 90, 105, 120, 135, 150	105
	vowel contrast	48, 63, 78, 93	97.5	105

Table I, each session included different subtests. See below for the procedure of each task. Training and identification test of natural speech materials took place in all sessions, while identification and discrimination of synthesized sounds took place on different days. Identification performance of new natural speech materials (transfer) was tested only on the last day. Throughout the sessions, no explicit explanation of the difference between the two categories was given to the participants.

1. Training

Each training session started with a label-learning task followed by an identification task. First, participants were presented with two example categories (/asu/-/assu/). The low-variability condition used the same example pair (spoken by F1) in all five training sessions while the high-variability condition used example pairs recorded by five different speakers (one speaker per session). The presentation of each example minimal pair was repeated six times with an ISI of 2000 ms. No response was required during the learning task. In the following identification task, one CVCV word was presented per trial and participants made categorical judgments. Feedback on the correctness of response was given as visual letters for 2000 ms after the response press. The ISI was set to 1500 ms. One training session consisted of 5 blocks of 32 trials, which took approximately 15 min in total.

2. Identification test/transfer test

The identification and transfer tests started with a brief category-label learning task followed by an identification task. During the labeling task, participants were presented with six repetitions of three example categories, e.g., /asu/-/assu/-/a:su/. No explicit explanation of the difference among the three categories was provided. 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 the five experiment 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 following 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 ISI was set to 1500 ms after the response key-press. The whole test took approximately 8 min. The transfer test employed the same procedure and consisted of 450 trials (90 trials per five transfer conditions).

3. Categorization of synthesized stimuli

The speech and nonspeech categorization tests followed the same procedure. The experiment started with a learning task followed by three blocks of an identification task. During the learning task, participants were presented with examples of two categories (singleton and geminate, or nonspeech analogs). The synthesized stimuli with the shortest (analog) C duration (60 ms) and the longest C duration (150 ms) combined with three vowel durations (43, 68, and 89 ms) were used as examples. These three minimal pairs were presented twice, which resulted in six presentations in total. Each category was presented along with its corresponding visual number with an ISI of 2000 ms.

During the categorization task, an auditory stimulus was presented and participants indicated their identification judgments. Presentation of the three continua (i.e., the short-, medium-, and long-vowel conditions, or their nonspeech analogs) was randomized in a single block. Each test included three blocks that each consisted from 42 trials, and lasted approximately 7 min.

4. Discrimination

The speech and nonspeech discrimination tests followed the same procedure. They employed a speeded 4I2AFC task. Each trial consisted of presentation of four stimuli. Either the second or the third stimulus of the four was a deviant. Participants were asked to indicate the position of that deviant by pressing the button "2" or "3." The probability of the deviant appearing in each position was set to 0.5. The test examined sensitivity to V_1 duration as well as C duration (or durations of the nonspeech analogs). The longest V_1 and C served as standard stimuli (V = 93 ms, C = 150 ms) and shorter durations served as deviants (see Table V for details). The ISI was set to 500 ms. The nine contrasts were randomly presented in a single block in order to keep the task challenging and to avoid boredom. There were five blocks, each consisting of 18 trials. Each test lasted approximately 9 min. Participants completed at least one practice session of six trials using a dummy word pair (/put/-/pet/; or a nonspeech analog) before the main session. No feedback was provided on task performance.

D. Apparatus

All sounds were recorded using a linear PCM recorder (Sony PCM–D1) with a sampling rate of 96 kHz in a sound attenuated booth. A DELL notebook computer with an IntellCoreDuo processer (4 GB RAM) was used to perform all experiments. Sony MDR-7506 headphones and a 15.4-in. TFT screen were used to present auditory stimuli and visual instructions, respectively. Average sound pressure level (SPL) of the headphones was adjusted to around 69 dB. The application presentation (version 14.3, Neurobehavioral Systems) was used for presenting instructions and stimuli as well as for collecting responses. The participants responded on the keyboard of the computer.

III. RESULTS

A. Training

Responses with a reaction time longer than three standard deviations from the grand mean (cutoff 2916.6 ms) were identified as outliers and excluded from the analysis (1.9% of all responses). Figure 1(a) shows the correct response rate in the first training session for the four types of training materials. A two-way analysis of variance (ANOVA) with block (continuous variable) and training material (LF1:low-variability

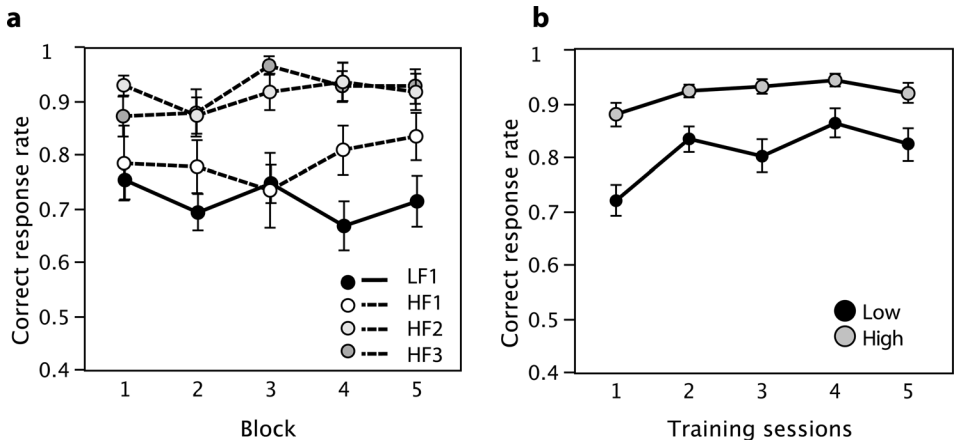

FIG. 1. (a) Correct response rate in the five blocks of the first training session and (b) over the five training sessions. Error bars indicate standard errors. LF1 = female voice 1 used for the low-variability training, HF1, HF2, HF3 = three female voices (F1, F2, F3) used for the high-variability training.

female voice 1, HF1:high-variability female voice 1, HF2:high-variability female voice 2, HF3:high-variability female voice 3) as independent variables and participants as random variable revealed a significant effect of training material [$F(3,26) = 9.0$, $p < 0.001$] with no significant effect of block [$F(1,116) < 1$, n.s.] and no interaction [$F(3,26) = 1.3$, n.s.]. Multiple comparisons confirmed that the correct response rate of LF1 was significantly lower than that of HF2 and HF3 ($p < 0.05$, all multiple comparisons are Bonferroni corrected). The correct response rate of HF1 did not significantly differ from any of the other three conditions.

Speaker F1, who was used for both testing and training, appeared to be the most difficult among the female speakers. Figure 1(a) indicates that participants had experienced difficulty performing the identification task in the first training session even with feedback. A control study, however, confirmed that native speakers of Japanese identified the geminate-singleton-long vowel contrast spoken by F1 fairly well without any training ($N = 10$, 91%), confirming the validity of this material. This control study tested only identification of the F1 materials.

Figure 1(b) shows the correct response rate for the five training sessions for the two groups. A two-way ANOVA with group (high/low-variability) and training session (a continuous variable) as independent variables and participants as random variable indicated effects of group [$F(1,28) = 24.3$, $p < 0.001$] and training session [$F(1,118) = 14.5$, $p < 0.001$] without a significant interaction [$F(1,118) = 2.6$, n.s.]. The 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 indicates that the high-variability group had a higher correct response rate.

B. Identification test

Responses with a reaction time longer than 3 standard deviations from the mean (cutoff 3646.2 ms) were identified as outliers and excluded (2.1% of responses). Figure 2(a) shows the identification accuracy in the pre-test (before the first training session) as well as in the five post-tests for the high- and low-variability groups. A t-test confirmed no significant difference between the groups' correct response rates in the pre-test [$t(28) = -1.59$, n.s.], indicating that performance of the two groups was equivalent before training.

Figure 2(b) shows improvement of identification accuracy relative to the pre-test. Relative improvement was calculated by subtracting the correct response rate of the pre-test from that of each of the five post-tests. A two-way ANOVA with training sessions (continuous variable) and group (high-/low-variability) as independent variables and participants as random variable indicated effects of training session [$F(1,118) = 79.7$, $p < 0.0001$] and group [$F(1,28) = 10.6$, $p < 0.001$] without a significant interaction [$F(1,118) < 1$, n.s.]. This indicates that both groups increased their response

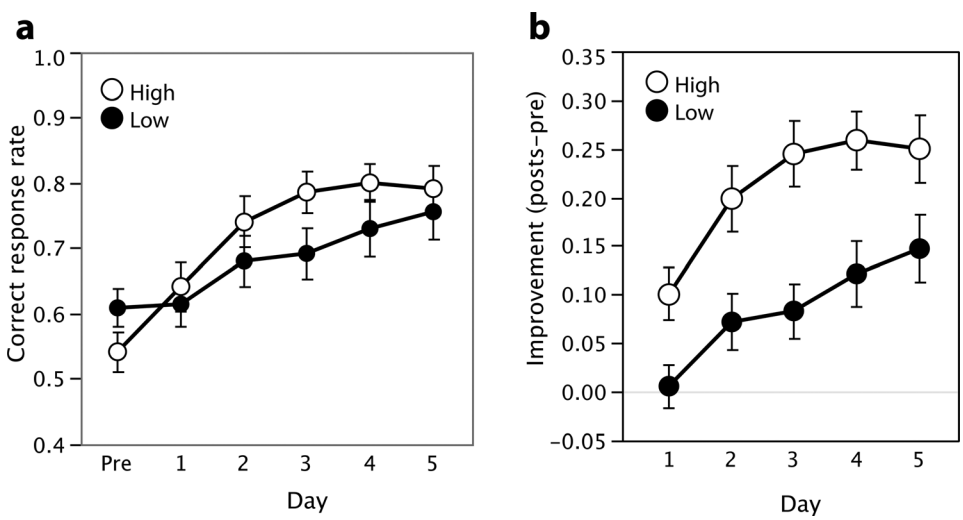

FIG. 2. (a) Correct response rate (absolute) and (b) Improvement of the response rate (relative). Error bars indicate standard errors. On the *x* axis of (a), "Pre" stands for the first pre-test; on the *x* axis of both panels, numbers stand for day of post-test.

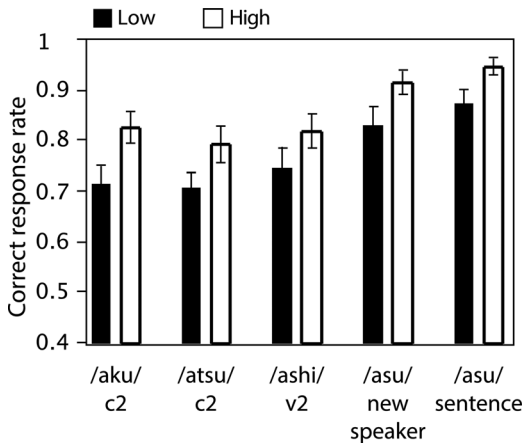

FIG. 3. Correct response rate in the five transfer conditions.

accuracy over the course of the five training sessions and that the participants in the high-variability condition had higher identification accuracy in all post-tests.

C. Transfer test

Due to a technical error, the transfer-test data of one participant in the high-variability group was lost; this participant was excluded from these analyses. A repeated measure ANOVA with group (high-/low-variability) as between-subject independent variable and condition (5 conditions) as within-subject independent variable indicated significant effects of group [$F(1,27) = 5.3$, $p < 0.05$] and condition [$F(4,24) = 18.5$, $p < 0.001$] without a significant interaction. As can be seen in Fig. 3, the high-variability group responded more accurately in all transfer conditions. *Post hoc* comparisons across the five conditions revealed in addition that response accuracy to the sentence and new speaker conditions was significantly higher than in the other three conditions ($p < 0.05$).

D. Discrimination test

Figure 4 compares response accuracy in the discrimination tests on days 1, 3, and 5 for the two groups as a function of the durational differences between the standard and deviant materials (durational differences for consonants or their nonspeech analogs: 15, 30, 45, 60, 75, and 90 ms; durational differences for vowels or their analogs: 15, 30, and 45 ms). The four plots show discrimination accuracy for speech and nonspeech stimuli for each group. Four mixed-model ANOVAs were performed with group (low-/high-variability) as between-subject factor and durational difference (six levels for consonant and three levels for vowel) and day (day 1/3/5) as within-subject factors. For all comparisons, no effect of group was observed (speech: consonant $F(1,28) < 1$, n.s., vowel $F(1,28) < 1$, n.s., nonspeech: consonant $F(1,28) < 1$, n.s., vowel $F(1,28) < 1$, n.s.). Type of training therefore did not significantly influence discrimination accuracy. The results for the speech stimuli indicated strong main effects of day for both consonant and vowel durations [consonant $F(2,56) = 9.85$, $p < 0.0001$; vowel $F(2,56) = 21.46$, $p < 0.0001$)]. Further analyses indicated that the correct

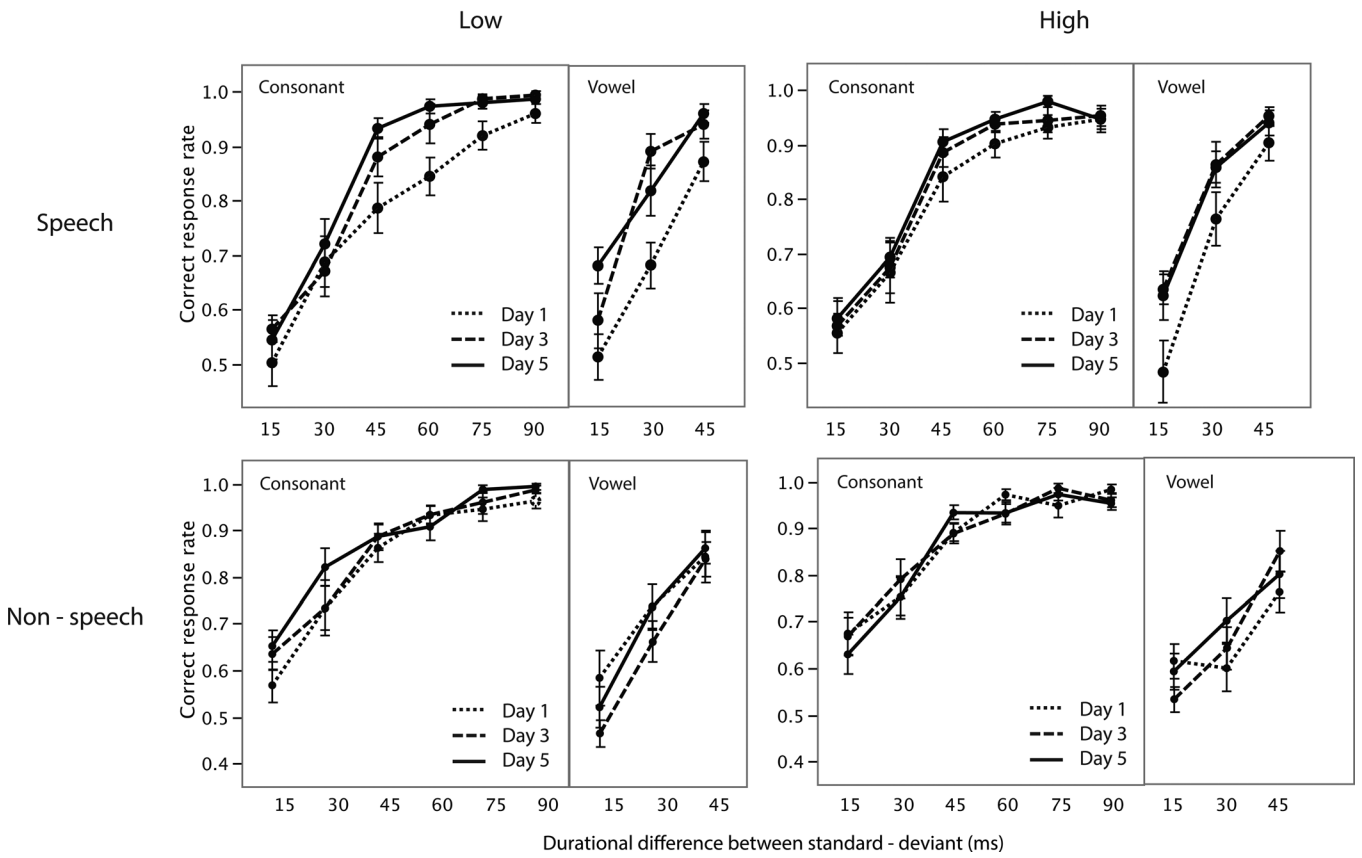

FIG. 4. Discrimination accuracy collapsed across the two groups, by day and type of material (speech vs nonspeech).

response rates on days 3 and 5 were significantly higher than on day 1 ($p < 0.05$), for consonants and vowels. There were also effects of durational difference [speech: consonant $F(5,140) = 183.39$, $p < 0.0001$, vowel $F(2,56) = 240.14$, $p < 0.0001$; nonspeech: consonant $F(5,140) = 104.81$, $p < 0.0001$, vowel $F(2,56) = 47.63$, $p < 0.0001$]. For the speech-consonant condition, larger improvements were observed for 45, 60, and 75 ms conditions, while larger improvements were observed at all levels of the speech-vowel condition. For the nonspeech stimuli, larger durational differences again resulted in higher discrimination accuracy. There was a significant interaction between duration difference and day for the nonspeech vowel condition. Further simple effect analyses revealed that the correct response rates of detecting differences of 150 and 300 ms were not significantly different on day 1, but became significantly different on days 3 and 5 ($p < 0.05$).

Before any training, individuals who showed higher accuracy in discriminating speech materials also showed higher accuracy when discriminating nonspeech materials ($r = 0.56$, $p < 0.01$). However, this correlation became weaker on day 3 ($r = 0.48$, $p < 0.01$) and nonsignificant on day 5 ($r = 0.21$, n.s.). This is because the training improved discrimination of the speech stimuli more than discrimination of the nonspeech stimuli.

E. Categorization of synthesized continua

Figure 5 shows percentage of "long" responses as a function of consonant duration in the three preceding vowel-duration conditions: short (48 ms), medium (63 ms), and long (93 ms), for days 1 and 5 and speech/nonspeech materials, collapsing over group (high-/low-variability). The slope of each participant's categorization function was estimated using logistic curve functions in PASW statistics (ver. 18). Larger coefficients reflect shallower slopes. Slope coefficients larger than 1.2 were treated as outliers because of expected poor estimation of nonlogistic data (Joanisse *et al.*, 2000). Table VI summarizes the mean slope coefficients and standard deviations for the speech and nonspeech materials in the three vowel conditions and the two sessions. These coefficients of speech and nonspeech materials were separately submitted to three-way mixed-model ANOVAs with group (high-/low-variability) as between-subject factor and vowel (duration of 48, 63, and 93 ms) and day (day 1/day 5) as within subject factors. The analysis of the speech data did

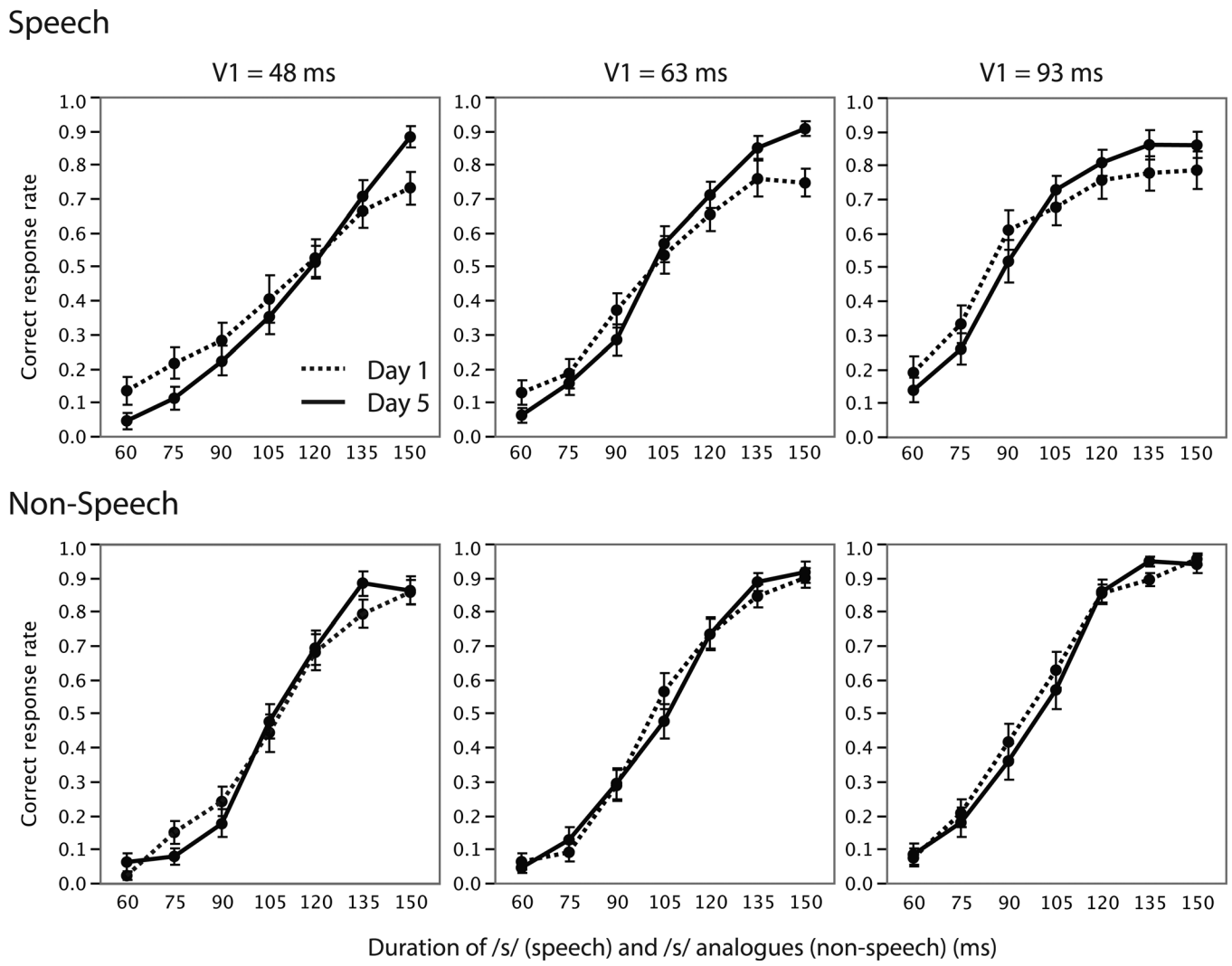

FIG. 5. Categorization of synthesized speech and nonspeech materials.

TABLE VI. Summary of slope coefficients of categorization functions in each condition.

		Speech		Nonspeech	
	Vowel duration	Day 1	Day 5	Day 1	Day 5
Low-variability	48	0.291 (0.25)	0.183 (0.23)	0.092 (0.11)	0.140 (0.27)
	63	0.218 (0.32)	0.174 (0.27)	0.133 (0.20)	0.110 (0.26)
	93	0.116 (0.16)	0.144 (0.25)	0.133 (0.19)	0.117 (0.26)
High-variability	48	0.192 (0.19)	0.090 (0.13)	0.146 (0.22)	0.166 (0.29)
	63	0.231 (0.25)	0.118 (0.22)	0.099 (0.17)	0.124 (0.23)
	93	0.173 (0.20)	0.060 (0.08)	0.048 (0.09)	0.084 (0.26)

not indicate a main effect of group [$F(1,23) < 1$, n.s.]. Thus, type of training did not significantly influence the steepness of the categorization function. An effect of day [$F(1,23) = 4.869$, $p < 0.05$] was indicated, suggesting that categorization functions were steeper for day 5 than day 1. A main effect of vowel [$F(2,46) = 3.833$, $p < 0.05$] and further *post hoc* comparisons revealed that slope coefficients were significantly smaller (thus steeper) in the long-vowel condition than in the short-vowel condition. The analysis of the nonspeech data did not indicate any significant effects.

F. Discrimination accuracy and perceptual learning

Figure 6 presents the improvement of identification accuracy (final post-test minus pre-test) as a function of first day overall discrimination accuracy on the speech materials. A simple linear regression analysis tested whether first-day discrimination accuracy (averaged over consonants/vowel) predicted relative improvement of identification accuracy (day 5 post-day 1 pre). There was a weak trend: [$t(28) = 2.59$, $\beta = 0.614$, $p = 0.06$]. Thus, individuals who showed higher discrimination accuracy (i.e., higher perceptual sensitivity) prior to training tended to improve more, but the effect was not significant.

IV. DISCUSSION

This study investigated the role of variability in perceptual learning of the Japanese geminate-singleton fricative consonant contrast by native listeners of Dutch. The high-variability training method, once again, was more effective for learning than the low-variability training method (replicating e.g., Logan *et al.*, 1991; Wang *et al.*, 1999; Hirata *et al.*, 2007; Tajima *et al.*, 2010). The benefit of the high-variability training method was already observable in the first post-test results (see Fig. 2), indicating that the bulk of this effect took place during the first training session. The benefit of high variability training was then stable over the remaining training sessions.

The high-variability group also performed fairly well in all conditions of the transfer test. It is perhaps not so surprising that the high-variability group generalized their training to a new vowel (/e/) and to a new speaker (M3) because these factors were varied in the high-variability materials. More strikingly, however, generalization was also demonstrated in enhanced performance accuracy of the high-variability group on the two new consonants (/k/, /ts/). Because the critical consonant /s/ was kept constant during the training, being able to perform the task well on these new consonant conditions requires generalization of learned phonological knowledge, that is, abstract knowledge that a consonantal distinction can be based on durational cues.

Further evidence that the high-variability participants had formed an abstract geminate / singleton contrast comes from the comparison of the discrimination and identification data. Accuracy in discriminating specific timing features in the speech materials increased as a result of training. Identification training thus enhances discrimination skills for relevant perceptual features. Importantly, however, low-variability training enhanced discrimination sensitivity just as much as high-variability training. This dissociation between tasks (a group difference for identification but not discrimination) is consistent with the view that, while the identification task taps into categorical representations, the 4I2AFC discrimination task taps into pre-categorical processing (Gerrits and Schouten, 2004; Sjerps *et al.*, 2013). This dissociation indicates that the high-variability group's superior accuracy in identifying natural speech materials was not because of their enhanced discrimination sensitivity to timing features, but rather because of their enhanced categorical representations. Variability in training materials thus helps nonnative listeners establish more robust abstract sound categories.

We varied the number of voices and word example pairs at the same time in order to maximally contrast the variability in the two sets of training materials. This makes it impossible to say which type of variability contributed more to the enhancement of learning. If speech rate counts more for learning Japanese geminates (Tajima *et al.*, 2010), then speaker variability rather than lexical variability is likely to have been more helpful over the entire experiment, because timing variation was larger among than within speakers.

Note that there was a delay in when learning started to have an effect in both variability groups. Even with feedback, virtually no learning occurred during the course of the first training session [see Fig. 1(a)]. Nevertheless, the high-variability group already started to show improvement at the post-test on day 1, while the low-variability group did not (see Fig. 2). Learning a less extreme durational difference (in the low-variability materials) seemed to require a longer period of time than learning a more variable and contrastive durational difference (in the high-variability materials). Intriguingly, both groups enjoyed large overnight improvements from

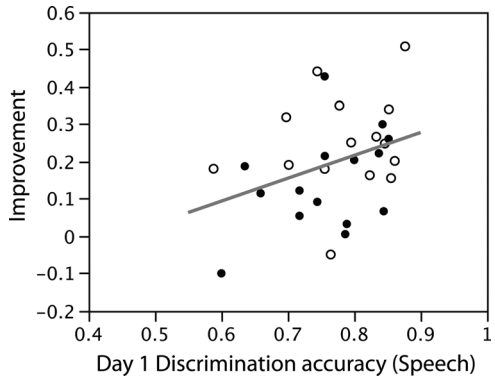

FIG. 6. Improvement of identification accuracy (final post-test minus pre-test) as a function of first day overall discrimination accuracy of speech stimuli (filled circles = low-variability group, open circles = high-variability group).

day 1 to day 2 (see Fig. 2). This effect may be related to consolidation of perceptual learning during sleep (Fenn *et al.*, 2003). This awaits further study.

The literature has repeatedly shown differences among individuals in their ability to learn to perceive L2 categories (e.g., Bradlow *et al.*, 1997; Hanulikova *et al.*, 2012; Strange and Dittmann, 1984). Our first research question, the role of variability, concerned a general environmental factor that may contribute to such differences. Other recent studies have looked into individual-specific factors that may also contribute to differences in perceptual learning (Hanulikova *et al.*, 2012; Perrachione *et al.*, 2011). Our second question related to one such individual-specific factor: Does an individual's discrimination accuracy prior to training predict improvements in identification accuracy? There was a trend suggesting that participants who showed higher discrimination sensitivity prior to training tended to improve more during the five sessions. Although the effect did not reach significance, it is reasonable to assume that higher sensitivity for relevant acoustic features (timing in this case) is useful when learning to identify categories. It would be interesting to investigate further whether there is an interaction between the listener's sensitivity and variability in timing-based category distinctions. Someone with higher sensitivity to durational information may benefit from variability but someone with lower sensitivity may suffer from variability when learning new sounds.

Our third research question concerned whether learning about geminates extends to nonspeech stimuli with similar durational properties. That is, is geminate training speech specific? There is much debate on possible associations between the mechanisms underlying the processing of linguistic and nonlinguistic information. Transfer of learning has indicated associations between acoustic information processing across domains. For example, musicians, who are extensively trained to deal with nonspeech sounds, often outperform nonmusicians when perceiving speech materials (Besson *et al.*, 2007; Sadakata and Sekiyama, 2011). An example of transfer in the reverse direction (speech to music) is a study showing that native tone-language speakers have more sensitivity than nontone-language speakers when perceiving pitch pattern information in nonspeech sounds (Pfordresher and Brown, 2009). Other studies, however, have found evidence that some aspects of perceptual learning are domain specific (Peretz and Coltheart, 2003; Peretz, 2009). For example, tone-language speakers show enhanced brainstem encoding of the pitch patterns of nonspeech materials (like musicians do), as compared to nontone-language speakers, but this enhanced response does not necessarily predict more accurate perceptual task performance (Bidelman *et al.*, 2011). Furthermore, Kingston *et al.* (2009) demonstrated that listeners use different strategies to analyze timing information in speech and nonspeech materials. In our study, the effect of perceptual learning was limited to speech materials: the training did not influence perception of nonspeech timing changes. This suggests that participants learned domain-specific rather than domain-general timing information. These results also indicate that the improvements observed for the speech discrimination tests are likely to reflect the specific enhancement of discrimination sensitivity of speech materials and are not an artifact caused by task familiarity.

Timing information is one of the most important acoustic cues for the Japanese geminate-singleton consonant contrast: differences are found for example in the length of the preceding vowel and the critical consonant (Kingston *et al.*, 2009) and in the ratio of closure to word duration (Hirata and Whiton, 2005; Amano and Hirata, 2010). Although previous studies have shown that learning to perceive this type of contrast is difficult for native speakers of English and German (Menning *et al.*, 2002; Motohashi-Saigo and Hardison, 2009; Hardison and Saigo, 2010; Tajima *et al.*, 2010), it was unknown whether this would also hold for native speakers of Dutch. Native listeners of Dutch make use of durational information in vowel identification (Smits *et al.*, 2003), in particular for the /a-a:/ contrast (Jongman *et al.*, 1992), although duration is not the only cue to this distinction (Booij, 1995). Dutch listeners may therefore be relatively good at dealing with timing-based cues in speech signals. Research involving other languages has indeed shown that sensitivity to temporal structure carries over from one's L1 to one's L2 in both production and perception (Bent *et al.*, 2008; Engstrand and Krull, 1994; Flege, 1993; Strange *et al.*, 1998). The pre-test in the present study indicated that the identification accuracy of the Dutch participants was not at ceiling, but this remained the case even after 5 days of training. Thus, the perceptual skills that native speakers of Dutch apply when perceiving vowel timing information apparently cannot be (easily or fully) generalized to consonants. This is in line with a previous finding of limited generalization of durational contrasts by nonnative speakers of Japanese: native speakers of English 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 (Tajima *et al.*, 2010).

V. CONCLUSION

The current study demonstrated that stimulus variability in training helps native speakers of Dutch to learn the

Japanese geminate-singleton contrast. Variability in training improved identification but did not enhance discrimination sensitivity, and this benefit held only for speech materials. Furthermore, variability in training led to better transfer of learning. Thus, while pre-categorical sensitivity to auditory signals is certainly involved in the process of identifying geminates and singletons, the benefit for nonnative listeners arising from high-variability training appears to arise at a domain-specific and categorical level of processing. Specifically, we suggest that this enhancement is because variability helps in the formation of an abstract geminate-singleton category contrast.

ACKNOWLEDGMENTS

This research was carried out while the first author worked at the Max Planck Institute for Psycholinguistics. The authors are grateful for support provided by Leonhaldt Rau and other research assistants at the Max Planck Institute for Psycholinguistics. We are also grateful for useful comments from Anne Cutler, Mirjam Broersma, and Holger Mitterer. Some of these results were presented at InterSpeech 2011.

Amano, S., and Hirata, Y. (**2010**). "Perception and production boundaries between single and geminate stops in Japanese," J. Acoust. Soc. Am. **128**, 2049–2058.

Bent, T., Bradlow, A. R., and Smith, B. L. (**2008**). "Production and perception of temporal patterns in native and non-native speech," Phonetica **65**, 131–147.

Besson, M., Schön, D., Moreno, S., Santos, A., and Magne, C. (**2007**). "Influence of musical expertise and musical training on pitch processing in music and language," Restor. Neurol. Neurosci. **25**, 399–410.

Bidelman, G. M., Gandour, J. T., and Krishnan, A. (**2011**). "Musicians and tone-language speakers share enhanced brainstem encoding but not perpetual benefits for musical pitch," Brain. Cogn. **77**, 1–10.

Boersma, P., and Weenink, D. (**2008**). "Praat: doing phonetics by computer (version 5.0.36) [computer program]," http//www.praat.org/ (Last viewed May 22, 2012).

Booij, G. (**1995**). *The Phonology of Dutch* (Clarendon, Oxford), pp. 13–16.

Bradlow, A. R., and Pisoni, D. B. (**1999**). "Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors," J. Acoust. Soc. Am. **106**, 2074–2085.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**, 2299–2310.

Engstrand, O., and Krull, D. (**1994**). "Durational correlates of quantity in Swedish, Finnish and Estonian cross-language evidence for a theory of adaptive dispersion," Phonetica, **51**, 80–91.

Fenn, K., Nusbaum, H. C., and Margollash, D. (**2003**). "Consolidation during sleep of perceptual learning of spoken language," Nature **425**, 614–616.

Flege, J. E. (**1993**). "Production and perception of a novel, 2nd language phonetic contrast," J. Acoust. Soc. Am. **93**, 1589–1608.

Gerrits, E., and Schouten, M. E. H. (**2004**). "Categorical perception depends on the discrimination task," Percept. Psychophys. **66**, 363–376.

Han, M. S. (**1992**). "The timing control of geminate and single stop consonants in Japanese: A challenge for nonnative speakers," Phonetica **49**, 102–127.

Hanulikova, A., Dediu, D., Fang, Z., Basnakova, J., and Huettig, F. (**2012**). "Individual differences in the acquisition of a complex L2 phonology: a training study," Lang. Learn. **62**, 79–109.

Hardison, D. M., and Saigo, M. M. (**2010**). "Development of perception of second language Japanese geminates: Role of duration, sonority, and segmentation strategy," Appl. Psycholinguist. **31**, 81–99.

Hirata, Y., Whitehurst, E., and Cullings, E. (**2007**). "Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates," J. Acoust. Soc. Am. **121**, 3837–3845.

Hirata, Y., and Whiton, J. (**2005**). "Effects of speaking rate on the single/geminate stop distinction in Japanese," J. Acoust. Soc. Am. **118**, 1647–1660.

Idemaru, K., and Guion, S. G. (**2008**). "Acoustic covariants of length contrast in Japanese stops," J. Int. Phonetic Assoc. **38**, 167–186.

Joanisse, M. F., Manis, F. R., Keating, P., and Seidenberg, M. (**2000**). "Language deficits in dyslexic children: speech perception, phonology, and morphology," J. Exp. Child. Psychol. **77**, 30–60.

Jongman, A., Sereno, J. A., Raaijmakers, M., and Lahiri, A. (**1992**). "The phonological representation of [voice] in speech perception," Lang. Speech **35**, 137–152.

Kingston, J., Kawahara, S., Chambless, D., Mash, D., and Brenner-Alsop, E. (**2009**). "Contextual effects on the perception of duration," J. Phonetics **37**, 297–320.

Logan, J. S., Liverly, S. E., and Pisoni, D. B. (**1991**). "Training Japanese listeners to identify English /r/ and /l/: A first report," J. Acoust. Soc. Am. **89**, 874–886.

Menning, H., Imaizumi, S., Zwitserlood, P., and Pantev, C. (**2002**). "Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language," Learn. Memory **9**, 253–267.

Motohashi-Saigo, M., and Hardison, D. M. (**2009**). "Acquisition of L2 Japanese geminates: Training with waveform displays," Lang. Learn. Technol. **13**, 29–47.

Peretz, I. (**2009**). "Music, language and modularity framed in action," Psychol. Belg. **49**, 157–175.

Peretz, I., and Coltheart, M. (**2003**). "Modularity of music processing," Nat. Neurosci. **6**, 688–691.

Perrachione, T. K., Lee, J., Ha, L. Y. Y., and Wong, P. C. M. (**2011**). "Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design," J. Acoust. Soc. Am. **130**, 461–472.

Pfordresher, P. Q., and Brown, S. (**2009**). "Enhanced production and perception of musical pitch in tone language speakers," Atten. Percept. Psychophys. **71**, 1385–1398.

Pisoni, D. B. (**1977**). "Identification and discrimination of relative onset time of 2 component tones - implications for voicing perception in stops," J. Acoust. Soc. Am. **61**, 1352–1361.

Sadakata, M., and Sekiyama, K. (**2011**). "Enhanced perception of various linguistic features by musicians: a cross-linguistic study," Acta Psychol. **138**, 1–10.

Sjerps, M. J., McQueen, J. M., and Mitterer, H. (**2013**). "Evidence for precategorical extrinsic vowel normalization," Atten. Percept. Psychophys. **75**(3), 576–587.

Smits, R., Warner, N., McQueen, J. M., and Cutler, A. (**2003**) "Unfolding of phonetic information over time: A database of Dutch diphone perception," J. Acoust. Soc. Am. **113**, 563–574.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., and Jenkins, J. J. (**1998**). "Perceptual assimilation of American English vowels by Japanese listeners," J. Phonetics **26**, 311–344.

Strange, W., and Dittmann, S. (**1984**). "Effects of Discrimination-Training on the Perception of (R-L) by Japanese Adults Learning-English," Percept. Psychophys. **36**, 131–145.

Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., and Munhall, K. G. (**2010**). "Training English listeners to perceive phonemic length contrasts in Japanese," J. Acoust. Soc. Am. **123**, 397–413.

Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. (**1999**). "Training American listeners to perceive Mandarin tones," J. Acoust. Soc. Am. **106**, 3649–3658.

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (**2007**). "Musical experience shapes human brainstem encoding of linguistic pitch patterns," Nat. Neurosci. **10**, 420–422.