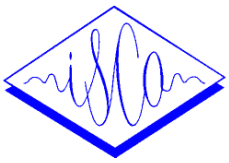
ISCA Archive
http://www.isca-speech.org/archive



2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

CROSS-LANGUAGES DIFFERENCES IN THE IDENTIFICATION OF INTERVOCALIC STOP CONSONANTS BY JAPANESE AND DUTCH LISTENERS

Makio Kashino[1], *Astrid van Wieringen*[2], *and Louis C. W. Pols*[2]

[1]NTT Basic Research Laboratories
Musashino-shi, Tokyo, 180 Japan

[2]Institute of Phonetic Sciences, University of Amsterdam
Herengracht 338, 1016 CG Amsterdam, The Netherlands

ABSTRACT

A joint Japanese and Dutch experiment studied the effect of language on the perceptual contribution of temporally distributed cues (i.e. VC transitions and CV transitions) in identifying intervocalic voiceless stop consonants. Japanese and Dutch listeners identified Japanese stop consonants from isolated VCV, CV, VC, or VC_1-C_2V syllables (all extracted from natural Japanese VCV utterances) under varying noise conditions, in which up to 70 ms of the release burst and the vocalic transition are replaced by noise. The two language groups perform very similar for the VCV and CV stimuli; the average consonant identification scores of both groups are higher for the VCV stimuli than for the CV stimuli when the CV cues are largely eliminated by noise replacement. The difference between the results of the two types of stimuli indicates that Japanese as well as Dutch listeners can use the VC cues (present in VCV stimuli) in identifying the stop consonants. However, the identification scores for the VC and VC_1-C_2V stimuli are lower for Japanese than for Dutch listeners, suggesting that the phonotactic system influences identification. Due to the absence of syllable-final stop consonants in the Japanese language, Japanese listeners are less inclined to use the VC transitions as cues to a syllable-final stop consonant than Dutch listeners.

10.21437/ICSLP.1992-322

I. INTRODUCTION

Several acoustic cues distributed over time, such as release bursts, closure duration, and spectral transitions, are integrated perceptually to cue stop consonants [1]. Examination of the relative importance of these acoustic properties in isolation or in combination [2, 3, 4, 5] has shown cues in combination to interact in a complex way in that they can compete, trade off, or cooperate with each other [6, 7, 8, 9, 10, 11].

However, very few studies have investigated the extent to which the perception of stop consonants depends on the language system. Fujimura et al. found Japanese and American English listeners to respond differently to the accent pattern of Japanese VCV stimuli, of which the VC and CV transitions cue different stop consonants [6]. American English listeners showed a greater tendency to identify the intervocalic consonant according to the VC transitions when the accent pattern was high-low than low-high. In contrast, Japanese listeners' judgments were unaffected by the accent pattern of the stimuli. Another Japanese study showed that Japanese listeners have great trouble in identifying unvoiced stop consonants from the VC transition only, due to the absence of syllable-final stop consonants in the Japanese language [10].

Since the perceptual integration of temporally distributed cues is to some extent dependent on the language system, the present study not only examines the relative importance of the formant transitions for stop consonant identification in isolation and in combination, but also the role of phonotactic constraints. This is done by comparing Japanese and Dutch listeners' perception to unvoiced Japanese stop consonants in four stimulus conditions: (1) naturally produced VCV (*VCV stimuli*; all cues are present), (2) the first vocalic portion of the VCV (*VC stimuli*; only the VC transitional cues are present), (3) the second vocalic portion of the VCV (*CV stimuli*; the release bursts and CV transitional cues are present), and (4) cross-spliced sequence of the first and the second vocalic portions from different VCV sequences (*VC_1-C_2V stimuli*; the VC cues and the CV cues signal different consonants). Contrary to the Dutch language, which has both syllable-initial and syllable-final unvoiced stop consonants, the Japanese language only contains syllable-initial stop consonants. Therefore, it is predicted that the perceptual role of VC transitions are different in the two languages.

II. EXPERIMENTS

2.1 Method

Stimuli. Both Japanese and Dutch subjects listened to the same Japanese stimuli. The stimulus conditions were a subset of those examined in [10]. For that study 45 disyllabic VCV nonsense sequences were recorded on a digital audio tape (DAT) by a male Japanese speaker (Tokyo dialect) with a low-high accent pattern. The speaking rate was moderate and the total duration of each utterance was approximately 500 ms. The recorded utterances were down-sampled at 16 KHz and transferred to a personal computer (NEC PC-9801ES2) through a digital audio interface (Iwatsu ISEL IS-3690), on which they were processed.

The first vowel was either /a/, /i/, /u/, /e/, or /o/, the intervocalic consonant was /p/, /t/, or /k/, and the second vowel was /a/, /e/, or /o/. Four conditions were tested, i.e. the original VCV, a pre-closure (VC), a post-closure (CV), and cross-spliced condition (VC_1-C_2V). This last subset was made by cross-splicing the first vocalic portion of a VC_1V sequence with the closure and the second vocalic portion of a VC_2V sequence. The vowel context of the two sequences was the same. For example, the first portion of /ate/ was spliced with the second portion of /ape/ resulting in a /at-pe/ stimulus. The stimuli sounded natural for both groups of listeners.

In order to suppress the release burst and formant transitions perceptually for stop consonant identification and to avoid artifacts caused by waveform truncation (silence replacement) [2, 11], five different quantities of white noises, i.e. 0 ms, 10 ms, 30 ms, 50 ms, and 70 ms, replaced the beginning of the vocalic portion in the CV stimuli, the beginning of the second vocalic portion in the VCV and VC_1-C_2V stimuli, or the end of the first vocalic portion in VC

stimuli. If sufficient perceptual cues remain in the un-replaced portion, phonemic restoration occurs and the appropriate phoneme is perceived [10, 11, 12].

The processed stimuli were up-sampled at 48 KHz and transferred to a DAT recorder (SONY DTC-500ES). In total, there were 225 items for the CV, VC, and VCV conditions and 450 items in the cross-spliced condition (VC_1-C_2V).

Subjects. Ten Japanese and ten Dutch native speakers participated in the phonetic identification tests. Contrary to the Dutch listeners, none of the Japanese listeners were familiar with listening tests. Four of the Japanese listeners could speak English well.

Procedure. The subjects were tested individually in sound-insulated booths (one in Japan and one in the Netherlands). The stimuli were presented diotically from the DAT through headphones at a comfortable loudness level in a random order (fixed on tape). The subjects were required to choose between either /p/, /t/, or /k/ in the CV, VC, and VCV conditions and between, either /pt/, /pk/, /tp/, /tk/, /kp/, or /kt/ for the VC_1-C_2V stimuli. No feedback was given. All four conditions were tested separately, preceded by a number of test trials. The Japanese listeners were given 5 minute breaks after every 100 trials. Stimuli were presented only once at interstimulus intervals of 4 s in all but the cross-spliced condition for Japanese subjects. As Japanese listeners were unable to respond after only one presentation in this condition, the VC_1-C_2V stimuli were repeated three times at intervals of 1 s. A pure tone(1000 Hz, 200 ms) was presented 1 s before the onset of each stimulus.

2.2 Results

VCV stimuli. **Table 1** shows percentages of perceived consonant for VCV stimuli as a function of noise duration, V_1 and V_2. Results for Japanese listeners and Dutch listeners are very similar. Even when the release bursts and CV transitions were entirely replaced by 70 ms of noise, more than 70% of the original consonants were correctly identified. Interesting enough, both Japanese and Dutch listeners show similar stop consonant confusions for similar vowel contexts. In both groups, the perception of /k/ is biased towards /p/ if the second vowel is an /o/ in context of all first vowels (39% for Japanese, 30% for Dutch). Presumably, the perception of the stop consonant is influenced by the rounding of the vowel.

CV stimuli. Results for CV stimuli are shown in **Table 2.** Once again, the identification scores of the Japanese and Dutch groups are very similar. Performance decreased as the noise duration increased. However, the identification scores remained above chance level (33%) under all noise conditions. Once again

Table 1. Percentages of perceived consonants for VCV stimuli as a function of noise duration, V1, and V2.

		Japanese subjects													**Dutch subjects**											
Original	Perceived	Noise duration (ms)					V1					V2			Noise duration (ms)					V1					V2		
consonant	consonant	0	10	30	50	70	a	i	u	e	o	a	e	o	0	10	30	50	70	a	i	u	e	o	a	e	o
p	**p**	**100**	**97**	**90**	**87**	**83**	**97**	**92**	**88**	**91**	**89**	**88**	**89**	**97**	**100**	**95**	**81**	**78**	**77**	**89**	**97**	**80**	**86**	**85**	**85**	**81**	**96**
p	t	0	2	6	9	7	3	6	5	9	1	3	11	0	0	3	6	8	9	6	3	5	13	1	8	8	0
p	k	0	1	4	5	10	1	2	7	0	10	8	0	3	0	2	13	14	13	5	0	15	1	14	7	11	3
t	p	0	0	0	4	10	2	2	4	4	2	0	1	7	0	0	2	5	9	1	5	1	7	1	1	1	8
t	**t**	**100**	**100**	**99**	**95**	**89**	**98**	**95**	**96**	**96**	**98**	**98**	**98**	**93**	**100**	**99**	**93**	**91**	**88**	**97**	**86**	**98**	**93**	**97**	**94**	**97**	**91**
t	k	0	0	1	1	1	0	3	0	0	0	1	1	0	0	1	5	4	3	2	9	1	0	1	5	2	1
k	p	0	11	22	24	36	9	17	29	19	19	11	6	39	0	4	15	19	18	8	13	12	10	13	2	1	30
k	t	1	0	2	3	8	1	11	1	1	1	5	3	0	0	0	3	7	10	3	7	6	1	3	8	2	2
k	**k**	**99**	**89**	**76**	**73**	**56**	**90**	**73**	**69**	**80**	**80**	**84**	**91**	**60**	**100**	**96**	**83**	**74**	**72**	**89**	**81**	**82**	**89**	**83**	**90**	**97**	**68**
Mean correct id.		**100**	**95**	**88**	**85**	**76**	**95**	**86**	**84**	**89**	**89**	**90**	**93**	**83**	**100**	**96**	**86**	**81**	**79**	**92**	**88**	**87**	**89**	**89**	**90**	**92**	**85**

Table 2. Percentages of perceived consonants for CV stimuli as a function of noise duration, V1, and V2.

		Japanese subjects													**Dutch subjects**											
Original	Perceived	Noise duration (ms)					V1					V2			Noise duration (ms)					V1					V2		
consonant	consonant	0	10	30	50	70	a	i	u	e	o	a	e	o	0	10	30	50	70	a	i	u	e	o	a	e	o
p	**p**	**99**	**91**	**60**	**57**	**55**	**70**	**71**	**68**	**68**	**85**	**70**	**57**	**90**	**100**	**88**	**52**	**41**	**33**	**65**	**57**	**60**	**55**	**74**	**57**	**48**	**82**
p	t	1	9	35	37	35	25	25	28	27	11	23	40	7	0	9	32	39	40	25	21	27	29	19	26	37	10
p	k	0	0	5	7	9	5	4	4	5	3	7	3	3	0	3	16	20	27	9	21	13	16	7	18	15	8
t	p	0	1	11	22	31	15	7	12	12	19	13	9	16	0	3	5	17	31	10	7	12	11	15	6	8	19
t	**t**	**100**	**97**	**85**	**63**	**55**	**81**	**75**	**86**	**81**	**77**	**81**	**77**	**82**	**100**	**93**	**77**	**58**	**48**	**75**	**73**	**79**	**78**	**72**	**71**	**76**	**79**
t	k	0	2	5	15	13	4	18	2	7	3	6	14	1	0	3	19	25	21	15	20	9	11	13	23	16	2
k	p	0	19	31	33	37	27	22	21	23	27	10	4	58	0	7	21	29	34	22	11	15	20	23	9	3	43
k	t	0	5	15	26	33	14	21	15	17	13	32	11	5	0	0	13	27	32	13	11	18	13	16	22	13	8
k	**k**	**100**	**76**	**53**	**41**	**30**	**59**	**57**	**64**	**61**	**61**	**58**	**85**	**38**	**100**	**93**	**66**	**45**	**34**	**65**	**78**	**67**	**67**	**61**	**70**	**84**	**49**
Mean correct id.		**100**	**88**	**66**	**54**	**47**	**70**	**68**	**73**	**70**	**74**	**70**	**73**	**70**	**100**	**92**	**65**	**48**	**38**	**68**	**70**	**68**	**67**	**69**	**66**	**69**	**70**

Table 3. Percentages of perceived consonants for VC stimuli as a function of noise duration, V1, and V2.

		Japanese subjects													**Dutch subjects**											
Original	Perceived	Noise duration (ms)					V1					V2			Noise duration (ms)					V1					V2		
consonant	consonant	0	10	30	50	70	a	i	u	e	o	a	e	o	0	10	30	50	70	a	i	u	e	o	a	e	o
p	**p**	**69**	**58**	**35**	**33**	**30**	**63**	**30**	**56**	**32**	**44**	**44**	**41**	**50**	**91**	**73**	**47**	**43**	**35**	**50**	**55**	**60**	**64**	**59**	**54**	**52**	**67**
p	t	19	27	46	41	47	29	43	21	52	34	35	38	34	3	13	24	28	35	23	24	23	20	13	21	24	16
p	k	12	15	19	27	23	9	27	23	16	21	21	20	16	7	13	29	29	30	27	21	17	16	28	25	23	17
t	p	18	24	19	27	29	33	11	43	15	14	23	21	26	12	10	14	17	15	13	24	5	23	3	12	12	17
t	**t**	**63**	**61**	**63**	**51**	**54**	**57**	**60**	**34**	**71**	**71**	**55**	**64**	**56**	**73**	**76**	**72**	**63**	**55**	**61**	**46**	**88**	**59**	**86**	**67**	**72**	**65**
t	k	18	15	19	21	17	10	28	22	13	15	22	14	17	14	14	14	19	30	25	30	7	18	11	21	16	8
k	p	29	32	28	24	35	31	17	47	17	37	24	32	33	24	26	28	32	29	13	32	51	15	29	24	27	32
k	t	25	30	49	51	42	37	51	23	47	40	43	39	37	7	12	19	27	43	17	27	23	25	16	25	20	20
k	**k**	**45**	**38**	**23**	**25**	**23**	**33**	**33**	**29**	**37**	**23**	**33**	**30**	**30**	**69**	**62**	**53**	**41**	**28**	**71**	**41**	**26**	**61**	**55**	**50**	**54**	**48**
Mean correct id.		**59**	**52**	**40**	**36**	**36**	**51**	**41**	**40**	**47**	**46**	**44**	**45**	**45**	**78**	**70**	**57**	**49**	**40**	**61**	**47**	**58**	**61**	**67**	**57**	**59**	**60**

Notes for Tables 1 ~ 4: For each language, the responses obtained from 10 subjects were averaged. The percentage for each noise duration were computed over both vowel conditions. To calculate the percentage for V1, 5 conditions for noise duration and 3 conditions for V2 were pooled together. To calculate the percentage for V2, noise duration and V1 were pooled together. The numbers in bold characters indicate the percentage of correct response.

the identification of the stop consonant /k/ was confused with that of /p/ by Japanese (58%) as well as Dutch (43%) listeners. Moreover, both groups tended to respond /t/ instead of /p/ when the (second) vowel was an /e/ (40% for Japanese and 37% for Dutch listeners).

VC stimuli. **Table 3** shows the results for VC stimuli. Consonant identification was better for Dutch than for Japanese listeners by approximately 20% when the noise was shorter than 70 ms. Even in the no-noise condition, the unvoiced stop consonants were not identified perfectly: 78% for Dutch listeners, and only 59% for Japanese listeners. The pre-closure portion does not contain sufficient perceptual cues to signal stop consonants. Although Japanese listeners often reported that they did not hear a consonant, performance remained above chance level in all noise conditions. Both the Japanese and Dutch subjects identified the /k/ more poorly than the other two stop consonants. Despite the absence of the second vowel in this condition, detailed vowel context analyses showed a clear response bias for /p/ instead of /k/ when the second vowel had been a /o/ by both Japanese (33%) and Dutch (32%) listeners.

VC_1-C_2V stimuli. **Table 4** shows percentages of perceived consonants as C_1 (consonant cued by the first portion) and C_2 (consonant cued by the second portion) for VC_1-C_2V stimuli. As a whole, Dutch listeners performed better than Japanese listeners by about 10 ~ 25%. They were able to identify the C_1 correctly more than 70% when the noise replacement in the second vocalic portion was less than 30 ms, and about 60% when the noise was longer. However, Japanese listeners identified C_1 only 53% at the no-noise condition, and performance fell to 35% in the 70 ms noise condition. Increasing the noise had a greater effect for the identification of C_2 than C_1 for both language groups. For Dutch listeners performance fell from 96% to 43% and for Japanese listeners from 81% to 29% across noise conditions. The task was very difficult for both language groups; Dutch listeners often

Table 4. Results for VC1-C2V stimuli as a function of noise duration, V1, and V2.
The upper and lower half of the table shows the percentage of perceived consonants as C1 and C2 respectively for 6 types of original stimuli.

		Japanese subjects													Dutch subjects											
Original	Perceived	Noise duration (ms)					V1					V2			Noise duration (ms)					V1					V2		
C1-C2	C1	0	10	30	50	70	a	i	u	e	o	a	e	o	0	10	30	50	70	a	i	u	e	o	a	e	o
p-t	**p**	**67**	**57**	**53**	**41**	**37**	**55**	**50**	**43**	**57**	**50**	**51**	**46**	**56**	**79**	**71**	**65**	**55**	**59**	**65**	**73**	**58**	**74**	**61**	**67**	**60**	**71**
p-t	t	18	23	32	33	31	29	27	27	29	25	26	32	25	19	22	21	19	21	15	11	16	9	17	15	17	8
p-t	k	15	19	15	26	32	15	23	31	14	25	23	22	19	2	7	15	26	19	21	16	26	17	22	18	23	20
p-k	**p**	**47**	**47**	**43**	**41**	**32**	**46**	**37**	**36**	**45**	**47**	**46**	**40**	**40**	**71**	**75**	**61**	**53**	**56**	**68**	**65**	**59**	**55**	**68**	**69**	**63**	**57**
p-k	t	34	33	30	35	37	29	39	45	31	26	33	34	35	25	21	19	27	21	19	22	20	33	20	20	26	22
p-k	k	19	19	27	24	31	25	24	19	25	27	21	26	24	4	5	20	19	23	13	13	21	12	12	12	10	21
t-p	p	19	21	26	29	33	27	21	27	23	30	29	26	22	3	14	23	17	30	21	21	14	21	11	19	24	10
t-p	**t**	**56**	**50**	**37**	**42**	**35**	**40**	**49**	**41**	**45**	**45**	**38**	**41**	**53**	**84**	**70**	**57**	**55**	**48**	**62**	**55**	**69**	**65**	**64**	**61**	**56**	**71**
t-p	k	25	29	37	29	32	33	30	31	33	25	33	33	25	13	16	21	27	22	17	24	17	15	25	20	20	19
t-k	p	27	34	27	29	27	35	21	29	27	33	26	29	32	27	23	23	21	11	17	29	13	31	14	22	21	20
t-k	**t**	**54**	**47**	**47**	**44**	**34**	**39**	**55**	**41**	**49**	**41**	**47**	**46**	**43**	**72**	**74**	**64**	**58**	**66**	**69**	**58**	**75**	**57**	**75**	**62**	**69**	**69**
t-k	k	19	19	25	27	39	26	24	29	25	25	28	25	25	1	3	13	21	23	14	13	12	11	11	16	10	11
k-p	p	21	21	27	25	29	29	25	23	19	27	25	20	28	3	11	23	23	21	12	22	20	13	13	21	15	12
k-p	t	38	35	26	33	34	31	40	39	33	24	36	31	32	34	22	20	27	29	21	40	38	11	23	24	28	27
k-p	**k**	**41**	**43**	**47**	**42**	**37**	**40**	**35**	**39**	**47**	**49**	**38**	**49**	**39**	**63**	**67**	**57**	**50**	**51**	**67**	**38**	**43**	**75**	**64**	**55**	**57**	**60**
k-t	p	28	36	32	33	31	35	27	40	27	31	33	29	34	38	26	27	31	21	26	36	43	16	22	29	23	34
k-t	t	19	18	25	23	33	25	27	25	24	17	26	23	22	4	12	25	20	20	11	25	18	16	11	18	20	11
k-t	**k**	**53**	**46**	**43**	**45**	**37**	**40**	**47**	**35**	**49**	**52**	**41**	**48**	**44**	**58**	**62**	**47**	**49**	**59**	**63**	**39**	**39**	**68**	**67**	**54**	**57**	**55**
Mean correct id.		**53**	**48**	**45**	**42**	**35**	**43**	**46**	**39**	**49**	**47**	**44**	**45**	**46**	**71**	**70**	**58**	**54**	**57**	**66**	**55**	**57**	**66**	**66**	**61**	**60**	**64**

		Japanese subjects													Dutch subjects											
Original	Perceived	Noise duration (ms)					V1					V2			Noise duration (ms)					V1					V2		
C1-C2	C2	0	10	30	50	70	a	i	u	e	o	a	e	o	0	10	30	50	70	a	i	u	e	o	a	e	o
p-t	p	11	21	29	47	53	37	32	29	33	29	26	41	29	1	3	9	22	21	8	8	16	10	13	9	13	11
p-t	**t**	**82**	**69**	**55**	**39**	**31**	**56**	**55**	**55**	**55**	**53**	**54**	**49**	**62**	**96**	**91**	**71**	**59**	**46**	**70**	**76**	**73**	**81**	**63**	**68**	**73**	**78**
p-t	k	7	10	16	15	17	7	13	15	12	17	20	10	9	3	6	21	19	33	22	16	11	9	24	24	14	11
p-k	p	11	20	35	48	61	39	40	23	43	31	24	32	50	2	3	19	27	25	10	13	21	21	10	10	6	30
p-k	t	9	8	13	15	17	12	11	14	10	13	13	14	10	2	2	12	20	27	8	13	15	16	11	18	13	7
p-k	**k**	**81**	**72**	**52**	**37**	**22**	**49**	**49**	**63**	**47**	**56**	**63**	**55**	**40**	**96**	**95**	**69**	**53**	**48**	**82**	**74**	**63**	**63**	**79**	**72**	**81**	**63**
t-p	**p**	**81**	**66**	**38**	**37**	**37**	**47**	**57**	**50**	**60**	**46**	**47**	**38**	**71**	**96**	**81**	**53**	**51**	**38**	**59**	**61**	**65**	**70**	**64**	**60**	**49**	**82**
t-p	t	8	19	43	47	46	39	21	37	25	41	33	49	16	2	9	27	29	41	28	17	21	17	25	25	33	7
t-p	k	11	15	19	16	17	15	22	13	15	13	20	13	13	2	10	20	20	21	13	21	15	13	11	15	18	11
t-k	p	7	17	20	25	29	17	15	21	30	15	11	12	36	1	1	23	30	39	15	15	21	27	17	9	5	43
t-k	t	11	13	28	40	49	28	20	32	23	38	35	28	22	1	3	12	26	19	11	7	14	17	13	20	11	6
t-k	**k**	**81**	**70**	**52**	**35**	**23**	**55**	**65**	**47**	**47**	**47**	**54**	**60**	**43**	**98**	**95**	**65**	**44**	**41**	**74**	**79**	**65**	**56**	**70**	**72**	**84**	**51**
k-p	**p**	**79**	**74**	**45**	**37**	**33**	**39**	**54**	**71**	**47**	**59**	**52**	**51**	**58**	**95**	**81**	**52**	**41**	**42**	**56**	**61**	**65**	**55**	**73**	**19**	**51**	**79**
k-p	t	10	13	30	23	21	28	20	15	21	13	14	32	12	5	11	27	31	30	30	20	13	31	11	56	30	8
k-p	k	11	13	25	40	46	33	26	15	33	28	34	17	29	0	8	21	28	28	14	19	21	15	16	25	20	12
k-t	p	10	9	15	20	28	13	15	22	15	18	15	17	17	1	5	11	15	30	5	7	17	13	19	9	12	14
k-t	**t**	**81**	**81**	**59**	**47**	**25**	**56**	**65**	**59**	**55**	**58**	**50**	**62**	**63**	**95**	**85**	**60**	**51**	**45**	**77**	**66**	**63**	**69**	**61**	**62**	**68**	**72**
k-t	k	9	11	25	33	47	31	21	19	31	24	34	20	20	4	10	29	34	25	17	27	21	18	19	29	20	14
Mean correct id.		**81**	**72**	**50**	**39**	**29**	**50**	**57**	**58**	**52**	**53**	**53**	**52**	**56**	**96**	**88**	**62**	**50**	**43**	**70**	**70**	**66**	**66**	**68**	**59**	**68**	**71**

Table 5. Percentages of cases where the stimuli were perceived as C1 or C2 under 3-alternative condition. For reference, results under 6-alternative conditions are also added. These numbers are drawn from the Mean correct id. column of Table 4 above.

	3AFC					6AFC				
Perceived	Noise duration (ms)					Noise duration (ms)				
consonant	0	10	30	50	70	0	10	30	50	70
C1	1	11	33	44	51	53	48	45	42	35
C2	98	84	54	40	30	81	72	50	39	29

reported that they could perceive only one consonant even in the no-noise condition, and Japanese listeners reported that they could hear no trace of C_1 throughout the session though they heard each stimulus three times.

III. DISCUSSION

Both Japanese and Dutch listeners could identify stop consonants in VCV stimuli better than in CV stimuli when the bursts and CV transitions were eliminated by the noise replacement. This means that the VC cues in the VCV stimuli actually did contribute to the perception of unvoiced stop consonants; when the most important cues in the second vocalic portion are replaced by noise, the VC cues improve the identification in cooperation with the cues occurring after the noise [10, 11]. As both Japanese and Dutch listeners are able to extract the VC transitions and use (or integrate) them as cues for a syllable-initial stop consonant both groups can be considered to be sensitive to the VC transitional cues.

However, Japanese listeners performed worse than Dutch listeners in identifying the syllable-final stop consonants from VC and VC_1-C_2V stimuli. Moreover, Japanese listeners' identification scores of C_2 in VC_1-C_2V stimuli were also worse than Dutch listeners' scores, despite C_2 being in the syllable-initial position. This result is in contrast with the previous studies in which Japanese subjects were asked to identify only one consonant (rather than two as in the present study) for the VC_1-C_2V stimulus [10, 11]. In those studies, Japanese subjects identified consonants according to the CV transitions (thus perceiving C_2) when the noise replacement in the C_2V segment was short, and as the noise duration increased, consonant identification gradually shifted to C_1, cued by the (undisturbed) VC transitions. An additional session, using a 3-alternatives forced-choice task and the 10 Japanese subjects who participated in the present study, exactly replicated the previous results (**Table 5**). When the noise replacement was shorter than 30 ms, identification of C_2 was significantly better in the 3AFC task than in the 6AFC task although the stimuli were identical. These results suggest that Japanese listeners have difficulty in processing stop-consonant clusters. For Japanese listeners, identifications of C_1 and C_2 are not mutually independent in the consonant clusters.

These results can be interpreted in terms of phonotactic constraints of the two languages. The Japanese language basically consists of consonant-vowel syllables, and has stop consonants only in the syllable-initial position. Therefore, in Japanese intervocalic stop consonants, acoustic events occurring before the stop closure are always related to the same stop consonant as those occurring after the closure. Contradictory to that, the Dutch language has syllable-final voiceless stop consonants as well as stop-consonant clusters. Therefore, for Dutch intervocalic stop consonants, acoustic events occurring before the closure may signal different consonants. Moreover, as the release burst is often absent at Dutch final stop consonants, Dutch listeners are used to listening more carefully to the VC than to the CV transitions. These cross-languages differences in phonological structure and its acoustic realization may affect the perceptual strategy in using multiple cues for stop consonants.

Undoubtedly, there are common aspects in stop consonant perception across different languages, too. For example, according to our vowel-context analyses, many confusions are similar for both the Japanese and Dutch language.

The process of integrating multiple cues for the perception of a phoneme is not well understood. Presumably, many types of constraints, ranging from those based on primitive auditory grouping to those based on speech-specific knowledge [1, 13] are involved. The present study has shown that perception of stop consonants can depend on the language system.

IV. CONCLUSION

The joint Japanese and Dutch experiment on stop consonant identification using identical stimuli and exactly the same procedure demonstrated the role of phonotactic constraints for the perception of multiple cues. The difference in identification scores of VCV and CV stimuli for the various noise replacements suggests Japanese listeners can use the VC transitions as a cue for a syllable-initial stop consonant as well as Dutch listeners. However, Dutch listeners performed better in identifying stop consonants from VC and VC_1-C_2V stimuli than Japanese listeners, suggesting that Japanese listeners have difficulty in using the VC transitions as a cue for a syllable-final stop consonant, due to the absence of such a segment in their language. The role of VC transitions in stop consonant perception depends on the languages.

REFERENCES

[1] B. H. Repp, "Integration and segregation in speech perception," *Language and Speech*, vol. 31, pp. 239-271, 1988.

[2] L. C. W. Pols and M. E. H. Schouten, "Identification of deleted consonants," *J. Acoust. Soc. Am.*, vol. 64, pp. 1333-1337, 1978.

[3] R. N. Ohde and D. J. Sharf, "Stop identification from vocalic transition plus vowel segments of CV and VC syllables: A follow-up study," *J. Acoust. Soc. Am.*, vol. 69, pp. 297-300, 1981.

[4] K. N. Stevens and S. E. Blumstein, "Invariant cues for place of articulation in stop consonants," *J. Acoust. Soc. Am.*, vol. 64, pp. 1358-1368, 1978.

[5] D. Kewley-Port, "Time-varying features as correlates of place of articulation in stop consonants," *J. Acoust. Soc. Am.*, vol. 73, pp. 322-335, 1983.

[6] O. Fujimura, M. J. Macchi, and L. A. Streeter, "Perception of stop consonants with conflicting transitional cues: A cross-linguistic study," *Language and Speech*, vol. 21, pp. 337-346, 1978.

[7] B. H. Repp, "Perceptual integration and differentiation of spectral cues for spectral cues for intervocalic stop consonants," *Percept. and Psychophys.*, vol. 24, pp. 471-485, 1978.

[8] M. E. H. Schouten and L. C. W. Pols, "Perception of plosive consonants: The relative contributions of bursts and vocalic transitions," in M. P. R. van den Broecke, V. J. van Heuven, and W. Zonneveld (Eds.) *Sound structures: studies for Antonie Cohen*, Foris Publication, Dordrecht, pp. 227-243, 1983.

[9] V. C. Tartter, D. Kat, A. G. Samuel, and B. H. Repp, "Perception of intervocalic stop consonants: The contributions of closure duration and formant transitions," *J. Acoust. Soc. Am.*, vol. 74, pp. 715-725, 1983.

[10] M. Kashino, "Distribution of perceptual cues for Japanese intervocalic stop consonants," *ICSLP90*, 14.1.1, pp. 557-560, 1990.

[11] M. Kashino, "Perception of Japanese intervocalic stop consonants based on the distributed cues in pre- and pre-closure portions," *J. Acoust. Soc. Jpn.*, vol. 48, pp. 76-86, 1992 (in Japanese).

[12] R. M. Warren, "Perceptual restoration of obliterated sounds," *Psychol. Bulletin*, vol. 96, pp. 371-383, 1984.

[13] A. S. Bregman, *Auditory Scene Analysis*, MIT Press, 1990.