Native language factors affecting use of vocalic cues to final consonant voicing in English

Court S. Crowther and Virginia Mann
Department of Cognitive Science, School of Social Science, University of California, Irvine, California 92717

(Received 28 June 1991; revised 4 March 1992; accepted 23 April 1992)

Preceding vocalic information can cue final consonant voicing for native English speakers. This study examines subjects' use of two vocalic cues, vocalic duration, and $F1$ offset frequency, as a function of two native language factors, experience with final stop consonants, and experience with phonemic vowel length. Native speakers of English are compared to native speakers of Japanese and Mandarin Chinese who are learning English as a second language. Experiment 1 measured the $F1$ offset frequency and vocalic duration in productions of "pod" and "pot." Experiment 2 assessed identification of natural tokens of "pod" and "pot" with and without closure segments and bursts. Experiment 3 assessed categorization of synthetic "pod"–"pot" stimuli that systematically manipulated vocalic duration and $F1$ offset frequency. Native speakers of English showed the strongest implementation of and sensitivity to vocalic duration, Mandarin speakers showed significantly weaker effects, and Japanese speakers fell in between. Group-related differences in use of $F1$ offset frequency were smaller, and much clearer in the case of production than perception. It is hypothesized that a lack of experience with final consonants penalizes use of vocalic cues among the Japanese and Mandarin subjects, in general. The advantage of the Japanese subjects for use of vocalic duration further suggests a facilitation effect of experience with phonemic vowel length.

PACS numbers: 43.71 Hw, 43.71 Es, 43.70.Kv

INTRODUCTION

It is generally agreed that for native English speakers, phonological voicing in final stop consonants is signaled primarily by the presence versus absence of glottal pulsing during the closure segment: Voiced stops are signaled by glottal pulsing, and voiceless stops by absence of glottal pulsing. Further, perceptual and acoustical studies of native American English speakers have indicated that a vocalic segment that precedes a consonant also contains spectral and temporal cues which influence the perception of consonant voicing. For native English speakers, these cues include, but are not limited to: vocalic segment duration (House, 1961; Raphael, 1972); $F1$ offset frequency (Fischer and Ohde, 1990; Summers, 1988; Walsh and Parker, 1983; Wolf, 1978; Hillenbrand *et al.*, 1984); and $F1$ steady state frequency (Wolf, 1978).

In this study, we have considered native speakers of languages that do not permit final stop consonants. We focus on their use of two of the more thoroughly researched vocalic cues to final consonant voicing that are used by native speakers of English: vocalic duration, and $F1$ offset frequency. The use of each of these two cues is examined in both production and perception of English CVC syllables by native speakers of American English, and by native speakers of Japanese and native speakers of Mandarin Chinese who have learned English as a second language. Japanese and Mandarin are chosen in light of two native language factors that may bear on use of vocalic cues in English CVCs: experience with final stop consonants and experience with phonemic vowel duration.

In contrast to English, Japanese and Mandarin Chinese share a phonotactic constraint on final stop consonants. Flege (1989) and Flege and Wang (1989) have suggested that if a speaker's native language does not include final stop consonants (and few final consonants, in general), then he or she may have difficulty using the vocalic cues to final consonant voicing in English. By comparing Japanese and Mandarin Chinese to native English speakers, we may clarify the influence of a lack of experience with final stop consonants on speakers's ability to implement vocalic duration and $F1$ offset frequency as cues to final consonant voicing.

While they share a constraint on final stop consonants, Japanese and Mandarin Chinese differ in another property which, as we shall argue in the following paragraphs, could also be relevant to use of the vocalic duration and $F1$ offset frequency cues in English. That difference concerns the durational contrast between phonemically long versus short vowels. The phonemic inventory of Japanese, unlike that of English or Mandarin Chinese, includes long versus short vowels, and the results of some recent work suggest that native experience with long/short vowels might be relevant to the use of vocalic duration as a cue to final consonant voicing (see, for example, Flege, 1984; Flege and Hillenbrand, 1986; Stevens *et al.*, 1986).

We have chosen to study vocalic duration because it is probably the most thoroughly studied vocalic cue to final consonant voicing, and because cross-linguistic studies suggest that implementation of vocalic duration as a voicing cue in both production and perception may be governed by language-specific factors. Chen (1970) reported that the vocalic segments produced by native speakers of Korean, English, French, and Russian were significantly longer when produced before voiced as compared to voiceless stop con-

 © 1992 Acoustical Society of America

sonants. However, there were large differences between language groups in the magnitude of the contrast, with native English speakers showing the greatest contrast, followed, respectively, by French, Russian, and Korean. Chen tentatively concluded that there is a universal trend toward producing longer vocalic segments before voiced as opposed to voiceless stop consonants, with unspecified factors governing the size of the effect. Chen's observation has been supported for many other languages (see Kluender *et al.*, 1988, for a compilation of these studies). However, more recent studies that have broadened the basis for comparison have challenged his contention about university (Flege and Port, 1981; Keating, 1985; Flege, 1988). These studies suggest that certain language-specific factors may influence the vocalic duration contrast.

For example, Flege and Port (1981) have reported that native speakers of Saudi Arabian Arabic, a language which includes phonemically long and short vowels, produced no significant durational contrast when producing Arabic CVCs and a significant vocalic duration contrast in only three of six English CVC minimal pairs (differing in final consonant voicing). Flege (1988) has also reported that native Mandarin Chinese speakers produce a relatively small vocalic duration contrast in English CVCs differing in final consonant voicing. The fact that native speakers of Mandarin show little use of vocalic duration is presumed to be a function of their lack of experience with final stop consonants (see Flege, 1989 and Flege and Wang, 1989). The fact that native speakers of Arabic do so as well, even though Arabic permits final stop consonants, would seem to suggest that the presence versus absence of final stops is not the only constraint on contrastive use of vocalic duration in production. The inclusion of long versus short vowels in one's native language may be another factor that influences the process of learning to use vocalic duration as a voicing cue in a second language.

There are also some indications that the syllabic context in which the voiced or voiceless consonant occurs should be considered by studies which examine the influence of voicing on vocalic duration. Chen (1970) compared the vocalic segment duration of CVCs from one language with CVCVs from another language, thereby confounding native language and syllabic context (Keating, 1985; and see Kluender *et al.*, 1988 for a review of additional difficulties with Chen's study). For example, although studies of CVCVs produced by native speakers of Japanese (Homma, 1981) indicate that they produce a small but significant vocalic duration contrast before voiced as opposed to voiceless stop consonants, another study of CVCV productions (Keating, 1985) reports no significant vocalic duration difference before voiced and voiceless stop consonants produced by native speakers of Polish and native speakers of Czech. Recognizing that the magnitude of the vocalic segment duration may depend upon syllabic context as well as language-specific factors, we have measured the vocalic duration contrasts produced by native speakers of English, Japanese, and Mandarin Chinese using the same set of CVC materials for each subject.

Another limitation in the current literature is that most researchers consider only production. Because use of vocalic duration in production may depend upon native language background, it is reasonable to ask whether native language also limits use of vocalic duration as a perceptual cue. At present, the answer is unclear. Although preceding vocalic duration has been shown to be an effective perceptual cue to final consonant voicing for native speakers of English (Raphael, 1972), there have been few assessments of the use of this cue by speakers of other languages. Consistent with the idea that native language factors may limit use of vocalic duration in both production and perception, Flege (1989) and Flege and Wang (1989) found native Mandarin speakers, who produce only a small vocalic duration contrast (Flege, 1988), to be comparatively insensitive to the vocalic duration cue (and possibly other vocalic cues) when listening to English CVCs. On the other hand, when Flege (1984) assessed the sensitivity of native English and native Arabic speakers to vocalic duration as a cue to final fricative voicing in English CVCs, he found that subjects in both groups were quite sensitive to the vocalic duration cue. Long vocalic segments encouraged perception of the voiced fricative /z/, while short vocalic segments encouraged perception of the voiceless fricative /s/ for the native Arabic speakers and the native English speakers, alike.

Flege's findings with native Arabic speakers raise the possibility that there may be no necessary correspondence between the effects of native language experience on production as opposed to perception. Native Arabic speakers seem to produce a negligible vocalic duration contrast when producing English or Arabic CVC tokens ending in voiced or voiceless stop consonants (Flege and Port, 1981), yet they are sensitive to vocalic duration as a cue to the voicing characteristic of English CVC tokens ending in voiced or voiceless fricatives (Flege, 1984). His findings for Mandarin Chinese speakers, however, imply that use of vocalic duration in perception parallels its use in production. There are at least two explanations for these apparent inconsistencies. One, offered by Flege (1984), appeals to the fact that Arabic includes a phonemic vowel length distinction. He suggests that, in his "peas–peace" perception experiment, the Arabic speakers may not have perceived the tokens in a native-English-like manner. Rather, they may have applied a post-perceptual strategy of identifying tokens with what they perceive to be phonemically long vocalic segments as "peas" and short vocalic segments as "peace." A second possibility is that different combinations of native language factors, such as final stop consonants and long versus short vowels, have different consequences on the ability of non-native English speakers to use vocalic dues to final consonant voicing when producing and perceiving English CVCs.

Aside from studying vocalic duration, we have also studied our subjects' use of $F1$ offset frequency as a cue to final consonant voicing. This should indicate whether any difficulties of non-native English speaking subjects are specific to vocalic duration as opposed to vocalic cues more generally. If, as Flege's hypothesis implies, a phonotactic constraint on final stop consonants penalizes the use of vocalic cues (Flege, 1989), then we should expect Mandarin Chinese and Japanese speakers to have difficulty with both

vocalic duration and $F1$ offset frequency. If the inclusion of a phonemic durational contrast between long and short vowels more directly bears on the use of vocalic duration, we should not necessarily expect it to influence $F1$ offset. To examine this and other possibilities, our study of native speakers of English, Japanese, and Mandarin compares their use of $F1$ offset frequency as a voicing cue in production and perception.

A number of studies assessing native English speakers' productions of CVC and VC syllables have shown that the $F1$ offset frequencies are lower when syllables end in voiced as compared to voiceless stop consonants (Wolf, 1977; Walsh and Parker, 1983; Hillenbrand *et al.*, 1984; Summers, 1987). However, the use of $F1$ offset in production remains to be examined for languages other than English. Research has likewise shown that, for native speakers of English, the effect of $F1$ offset frequency in perception is consistent with $F1$ offset frequency in production: Relatively low $F1$ offset frequencies favor voiced percepts, while relatively high $F1$ offset frequencies favor voiceless percepts (Wolf, 1977; Walsh and Parker, 1983; Hillenbrand *et al.*, 1984; Summers, 1988; Fischer and Ohde, 1990). However, it remains to be determined how native language background affects use of $F1$ offset frequency.

Our study is divided into three parts, each of which concerns vocalic duration and $F1$ offset frequency as cues to the distinction between /pod/ and /pot/. We have used a low vowel because $F1$ offset frequency as a final consonant voicing cue is generally more effective for low vowels than for high vowels (Hillenbrand *et al.*, 1984; Fischer and Ohde, 1990). Using higher vowels might not reveal any interesting between-groups differences for the simple reason that, with such vowels, even native English speakers may fail to use $F1$ offset to a very large extent as perceptual voicing cue. We have further chosen to confine our study to the /t/ vs /d/ distinction because many of the cross-linguistic studies of vocalic duration as a voicing cue have focused on this distinction (e.g., Keating, 1985; Flege, 1989; and Flege and Wang, 1989). Studying this place of articulation would thus enable us to more reasonably compare our findings with previous findings.

In experiment 1, we measure the vocalic duration and $F1$ offset frequency of subjects' productions of "pod" and "pot." In experiment 2, we study the perception of natural tokens of these syllables, extending Flege's work (1989; and Flege and Wang, 1989). A comparison is made between perception of unedited tokens and edited tokens from which the release burst and closure interval have been removed in order to test perceptual sensitivity to vocalic cues in natural speech. Experiment 3 focuses on perceptual use of vocalic duration and $F1$ offset frequency by orthogonally manipulating each of these cues and asking subjects to make a forced choice identification. The experiment uses synthetic speech stimuli in three different "pod–pot" continua which systematically vary vocalic duration and $F1$ offset frequency.

TABLE I. Relevant subject attributes. [SUB = subject; NL = native language; GEN = gender; AOA = age of arrival in US; and TUS = time (in months) spent in US.]

SUB	NL	GEN	AOA	TUS
HK	Japanese	F	20	0.5
NS	Japanese	M	66	1.0
AN	Japanese	F	20	0.5
MK	Japanese	F	21	0.25
MA	Japanese	F	20	0.5
MH	Japanese	F	25	0.5
KK	Japanese	F	21	0.5
TT	Japanese	M	26	5.0
HS	Japanese	M	27	0.75
YO	Japanese	M	25	0.25
ZW	Mandarin	M	25	24.0
HL	Mandarin	M	26	60.0
AW	Mandarin	F	24	36.0
ZM	Mandarin	M	24	60.0
EK	Mandarin	F	22	6.0
WH	Mandarin	M	25	12.0
LC	Mandarin	M	24	18.0
PS	Mandarin	F	25	18.0
CL	Mandarin	M	21	54.0
WW	Mandarin	F	18	1.0

I. EXPERIMENT 1: VOCALIC DURATION AND F1 OFFSET FREQUENCY IN PRODUCTION

A. Method

1. Subjects

Three groups of subjects participated in experiment 1 (see Table I for relevant subject characteristics). All subjects reported having normal hearing and, in the opinion the first author, did not exhibit obvious speech production problems. The native American English subjects (three females and seven males), were undergraduate students at the University of California, Irvine, and ranged in age from 18–22 years. They were compensated with extra credit in social sciences courses.

The Japanese subjects (six females and four males) were paid volunteers, drawn from a pool of Japanese nationals taking summer courses to learn English as a second language through the extension program at University of California, Irvine. As is common practice in Japan, each of the subjects had begun learning English during the Japanese equivalent of American junior high school. Four of them had native English-speaking tutors/teachers for more than 6 months. Half of them reported having listened to English educational TV/radio programs employing native English speakers for more than six months before coming to the US. Their mean duration of residency in the US was less than 1 month.

The Mandarin Chinese subjects (four females and six males) were paid volunteers, drawn from a pool of Mainland Chinese nationals, who were either attending school in the US or accompanying as a spouse. Each of the subjects had begun learning English during the Chinese equivalent of American junior high school. One of these subjects had had native English-speaking tutors/teachers for more than 6 months. Eight of them reported having listened to English educational TV/radio programs employing native English speakers for more than 6 months before coming to the US. Their mean duration of residency in the US was greater than 2 years.

2. Materials and procedure

Productions of /pod/ and /pot/ were elicited by having subjects read a fixed random sequence of sentences which consisted of six tokens of each of these words embedded in the carrier phase "Say—Chuck." The test sentences were randomly interspersed with sentences containing other CVC syllables: bib, bip; bag, back; peas, peace.[1] The subjects were told only to read each sentence; they were not instructed to release the final stop. Their utterances were recorded in a sound attenuating booth, using a Tascam tape recorder at 7 ½ ips and a Sony clip-on microphone. For the purpose of measuring vocalic duration and $F1$ offset frequency, the test sentences were digitized at 10 kHz and low-pass filtered at 4.9 kHz. Vocalic segment duration measurements were made directly from waveform representations. Following Summers (1987) and Mack (1982), the vocalic segment was defined as beginning at the onset of voicing (after the /p/ was released) and ending at the beginning of the closure segment. $F1$ offset frequency measurements were then made, using a formant tracking algorithm (with a 20-ms Hamming window) that was part of the Interactive Laboratory Systems (ILS) software system. Following Hillenbrand *et al.* (1984), the offset of $F1$ was indicated by a decrease in formant peak amplitude and "the presence of abrupt discontinuities with previously extracted formant frequencies" (p. 23).

B. Results and discussion

1. Preceding vocalic duration

Mean durations of the /pod/ and /pot/ productions are shown separately for the three groups in Table II. The mean difference in vocalic segment duration of /pod/ and /pot/ was greatest for native English speakers and considerably less for both Japanese and Mandarin Chinese speakers.

Our primary interest was in comparing the magnitude of the vocalic duration contrast produced by speakers in each of the three language groups. To make this comparison, we computed the mean magnitude of the vocalic duration contrast (mean vocalic duration for the six "pod" productions minus mean vocalic duration for the six "pot" productions) for each subject in each group. The average vocalic duration contrasts were 75.9, 25.5, and 12.8 ms for the native English, Japanese, and Mandarin Chinese speakers, respectively. The mean vocalic duration contrasts were then submitted to a one-way ANOVA with native language as the grouping factor. The group main effect was highly significant [$F(2,27) = 73.7$, $p \ll 0.001$]. *Post hoc* Newman–Keuls tests ($p = 0.05$) revealed that the vocalic duration contrast was significantly greater for the English than for the Japanese speakers; significantly greater for the English than for the Mandarin speakers; and significantly greater for the Japanese than for the Mandarin speakers. The 95% confidence interval for the vocalic duration contrast was 66.1 to 85.8 ms for the English group, 15.0 to 36.0 ms for the Japanese group, and 7.7 to 17.9 ms for the Mandarin group. This implies that the absolute duration contrast was significantly greater than zero for all three groups.

TABLE II. Mean vocalic duration (in ms) as a function of native language group and word. Standard deviations in parentheses.

Group	Word: pod	Word: pot
American English	198.9 (25.7)	123.0 (27.3)
Japanese	141.4 (33.8)	116.0 (23.1)
Mandarin Chinese	140.4 (17.5)	127.6 (15.7)

The vocalic duration results of experiment 1 are consistent with Chen (1970): There was a universal trend insofar as subjects in each group produced significantly longer vocalic segments before voiced as compared to voiceless stop consonants, but there were group differences in the magnitude of the effect. Native English speakers produced durational contrasts that were significantly greater than either native Japanese or native Mandarin speakers, and the contrast produced by Japanese speakers was significantly greater than that produced by the Mandarin speakers. Native experience with final stop consonants can explain the fact that the native speakers of English showed the greatest durational contrast. However, the difference observed between native Japanese and Mandarin speakers is probably not due to native experience with final stop consonants, because neither Japanese nor Mandarin allows stop consonants in final position. Also, experience speaking and listening to English may not account for the observed differences, because the Mandarin speakers, on average, had at least as much ESL training and had spent at least as much time in the US as the Japanese speakers (see Table I). Here phonemic inventory may be relevant; that the Japanese subjects tended to produce a greater vocalic duration contrast than the Mandarin subjects could be due to their experience with phonemically long and short vowels. As noted previously, unlike Mandarin Chinese, Japanese is a quantity language that includes phonemically long versus short vowels. It may be the case that native speakers of Japanese are able to apply their native language experience with long/short vowels, which are realized in the acoustic signal by a vocalic duration contrast, to the task of producing longer vocalic segments before voiced, as opposed to voiceless stop consonants when speaking English. This transfer of training argument could explain the performance difference between the Japanese and Mandarin speakers, although it is at odds with Flege and Port's (1981) observations about native speakers of Arabic. We shall return to this point in the final discussion.

2. F1 offset frequency

The mean $F1$ offset frequencies for "pod" and "pot" productions for subjects in each group are given in Table III. We were primarily interested in between-group differences in the magnitude of the $F1$ offset frequency contrast for subjects' "pot" as compared to "pod" tokens. To make this comparison, we computed the mean $F1$ offset frequency contrast (mean $F1$ offset frequency for "pot" tokens minus mean $F1$ offset frequency for "pod" tokens) for each subject in each group. The mean $F1$ offset frequency contrasts were 304,

TABLE III. Mean $F1$ offset frequency (Hz) as a function of native language group and word. Standard deviations in parentheses.

Group	Word	
	pod	pot
American English	293.5	597.2
	(64.2)	(82.5)
Japanese	371.9	473.2
	(89.2)	(112.6)
Mandarin	482.7	538.0
	(114.2)	(105.4)

101, and 55 Hz for the native English, Japanese, and Mandarin Chinese speakers, respectively. The mean $F1$ offset frequency value contrast were then submitted to a one-way ANOVA with native language as the grouping factor. The group main effect was highly significant [$F(2,27) = 26.8$, $p \ll 0.001$]. *Post hoc* Newman–Keuls tests ($p = 0.05$) revealed that the $F1$ offset frequency contrast was significantly greater for the English than for the Japanese speakers, and significantly greater for the English than for the Mandarin speakers. The contrast for the Japanese speakers was not significantly different from that of the Mandarin speakers. The 95% confidence interval for the $F1$ offset frequency contrast was 232 to 375 Hz for the English group, 33 to 170 Hz for the Japanese group, and 42 to 68 Hz for the Mandarin group. This implies that the absolute $F1$ offset contrast was significantly greater than zero for all three groups.

To review, subjects in each group produced significantly higher $F1$ offsets before voiceless as opposed to voiced stop consonants. However, subjects in the native English group produced a greater $F1$ offset frequency contrast than subjects in either of the non-native groups. The fact that the $F1$ offset difference for native speakers of English surpassed that of the native Japanese and Mandarin subjects could be due to experience with final stop consonants. The similarity of the Japanese and Mandarin data is consistent with the idea that experience with long versus short vowels is not a factor relevant to the use of $F1$ offset frequency in production.

The common trend to produce vocalic segments with higher $F1$ offset frequencies before voiceless stops could result from constraints imposed by the human vocal tract. For example, execution of the closure gesture for a voiced stop consonant may necessarily result in relatively low $F1$ offset frequencies for vocalic segments produced before voiced stop consonants. On the other hand, it may be the case that $F1$ offset contrast in production is not a necessary consequence human vocal tract anatomy. If $F1$ offset frequency can be controlled independently of final consonant voicing, then perhaps it is a language-specific phenomenon, in that some language communities may have chosen to exploit the perceptual salience of $F1$ offset frequency differences by "voluntarily" implementing the contrast in production to convey final consonant voicing information. This later possibility would be supported were we to find a language community in which low $F1$ offsets were correlated with voiceless, as opposed to voiced final stop consonants. More cross-linguistic production data will be needed to help settle this issue.

II. EXPERIMENT 2: PERCEPTUAL INFLUENCE OF VOCALIC CUES IN NATURAL CVC TOKENS

A. Introduction

Experiment 2 turns to the perceptual use of vocalic cues to final consonant voicing. In it we asked subjects to label unedited and edited natural tokens of /pod/ and /pot/. Recent work has suggested that the perceptual use of vocalic cues is mediated by phonotactic constraints imposed by a speaker's native language. Two of the more thorough cross-linguistic investigations of the influence of preceding vocalic segments on the perception of final stop consonant voicing studied native speakers of several different Chinese dialects. Flege (1989) and Flege and Wang (1989) reported that, although American English and Chinese subjects are highly accurate ($> 90\%$) in identifying naturally produced CVC minimal pair tokens differing in final consonant voicing, such as "bead" and "beat," the performance of certain native Chinese speakers is markedly degraded when these tokens are edited by removing the burst and closure intervals. When burst and closure are removed, those subjects unfamiliar with final stop consonants (native Mandarin speakers) make less effective use of the vocalic cues than those who are familiar (native speakers of Shanghainese and Cantonese).

Flege (1988) and Flege and Wang (1989) suggest that inclusion of final stop consonants is the property of a speaker's native language which determines the ability to use final consonant voicing cues which are present in the vocalic segment of CVC syllables. More specifically, if a speaker's native language includes final stop consonants, then he or she will have learned to allocate attention to syllable offset properties. Consequently, the degree to which a speaker is able to use vocalic cues in his or her native language, and when using a second language, is positively correlated with the number of final stop consonants included in his or her native language.

Experiment 2 is an extension of Flege's (1989) Experiment 1 to native speakers of American English, Japanese, and Mandarin Chinese. Flege's hypothesis predicts that native speakers of Japanese should behave more-or-less like Mandarin Chinese speakers because Japanese, like Mandarin, allows no final stop consonants. In addition, Japanese allows only /N/ in word-final position, and Mandarin allows only the consonants /m/ and/ng/ in word-final position. Lacking diverse native language experience with final consonants and with stop consonants, in particular, both Japanese and Mandarin Chinese subjects should be more penalized by editing than the native speakers of English.

B. Method

1. Subjects

The subjects were the same 30 English, Japanese, and Mandarin Chinese speakers who participated in experiment 1.

2. Stimuli

Each of two native American English speakers read, in citation form, three tokens of "pod" and three tokens of "pot" imbedded in a randomized list of eight different English CVC syllables. All final stop consonants were released. The tokens were digitized at 10 kHz and low-pass filtered at 4.9 kHz. For the six "pod" tokens, averaged across the two speakers, the mean vocalic duration was 205.7 ms (s.d. = 23.1); mean closure duration was 104.6 ms (s.d. = 23.3); and mean $F1$ offset frequency was 288 Hz (s.d. = 88). The mean rms amplitude for the vocalic segments was 449.3 mV (s.d. = 44.9), and the mean rms amplitude for the bursts was 42.8 mV (s.d. = 11.0). For the six "pot" tokens, averaged across the two speakers, the mean vocalic duration was 132.0 ms (s.d. = 20.1); mean closure duration was 133.0 ms (s.d. = 46.6); and mean $F1$ offset frequency was 590 Hz (s.d. = 96). The mean rms amplitude for the vocalic segments was 503.6 mV (s.d. = 56.6), and the mean rms amplitude for the bursts was 69.0 mV (s.d. = 16.9). In the "unedited" condition, each token was used in its entirety (i.e., initial burst, vocalic segment, closure interval, and release burst were unaltered); for the edited condition, all energy in the closure interval and final burst was zeroed. Stimuli were stored on the hard disk of a PC (IBM AT) and low-pass filtered at 4.9 kHz prior to presentation.

3. Procedure

Stimuli were presented to individual subjects with the "Identify" software system (Shannon *et al.*, 1988) at 75 dB via TDH-39 earphones. To familiarize the subjects with task, at the beginning of the experiment, subjects received a brief training session with response feedback. In the training, one unedited token of "pod" and one unedited token of "pot" were presented ten times each, in random order. The training process was repeated until the subject attained an accuracy rate of at least 75%. In the test that followed, the 12 tokens (two words × two speakers × three tokens) were randomized and presented without replacement in ten blocks (i.e., each token was presented ten times). The unedited and edited conditions were presented separately, and subjects always heard the unedited before the edited condition. Subjects were instructed to identify each token as either "pod" or "pot" by pressing a button labeled as "pod" or as "pot," and they were told to guess if unsure. Subjects' responses were recorded, by computer, with the Identify software. Shortly before each token was presented, subjects were cued visually (on the PC monitor) and aurally with the word "ready." Presentation of this word was contingent on a response to the previous stimulus, and followed the subject's response by 1 s.

C. Results and discussion

Figure 1 shows the mean percentage of tokens correctly identified as a function of subject group and editing condition, collapsed across the word variable. For analysis, the percent accuracy data were collapsed over the word variable (pod or pot), transformed using the arcsine function,[2] and

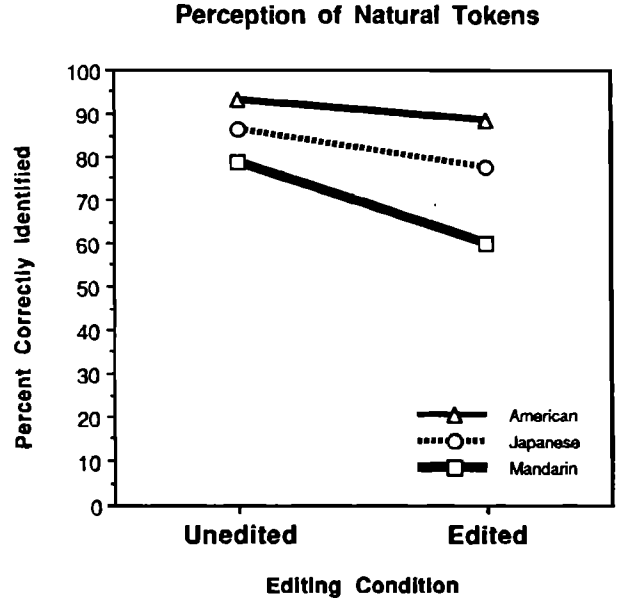


FIG. 1. Mean percentage of naturally produced "pod" and "pot" tokens correctly identified as a function of editing condition and subjects' native language background. In the edited condition, bursts and closure intervals were removed.

then submitted to a repeated-measures group (native English, Japanese, or Mandarin listener) by condition (unedited/edited) ANOVA. Main effects of group [$F(2,27) = 14.9$, $p \ll 0.01$] and editing condition [$F(1,27) = 37.5$, $p \ll 0.01$] were significant. The group by editing condition interaction [$F(2,27) = 1.7$, $p > 0.05$] was nonsignificant. To compare subjects' overall accuracy, the data were collapsed over the editing condition variable to test the simple main effect of group [$F(2,27) = 14.3$, $p \ll 0.01$]. *Post hoc* Newman–Keuls ($p = 0.05$) tests showed that the English speakers were significantly more accurate than the Japanese and the Mandarin Chinese speakers, and the Japanese speakers were significantly more accurate than the Mandarin speakers.

Flege's hypothesis predicts that both Japanese and Mandarin subjects should be markedly penalized by editing, but it does not predict a performance difference between native speakers of Japanese and native speakers of Mandarin Chinese, because neither Japanese nor Mandarin Chinese includes final stop consonants. We find that the performance of subjects in each group was degraded by editing, but we did not find the group by editing condition interaction to be significant. We cannot conclude that editing affected subjects differently, according to their language group membership; however, there is a nonsignificant trend in the data which suggests that editing degraded performance of subjects in the Mandarin group to a greater extent than it degraded the performance of subjects in the Japanese group. On average, editing resulted in a 8% decrease in performance for Japanese subjects, while it resulted in a 19% decrease for the Mandarin subjects. It may be the case that subjects in the Japanese group, as compared to subjects in the Mandarin group, made more effective use of one or more vocalic cues. Experiment 3 is designed to probe the use of vocalic duration and $F1$ offset frequency more directly.

III. EXPERIMENT 3: PERCEPTUAL INFLUENCE OF VOCALIC DURATION AND F1 OFFSET FREQUENCY IN SYNTHETIC TOKENS

A. Introduction

The observed performance differences in experiment 2 cannot be conclusively attributed to sensitivity to either $F1$ offset frequency or vocalic duration; in the edited condition, the removal of the closure interval and release burst forces the listener to rely on any available cues to final consonant voicing that are present in the vocalic segments, and degraded performance would be consistent with an inability to use either of these vocalic dues. Experiment 3 is intended to more closely examine perceptual sensitivity to the vocalic duration and $F1$ offset frequency cues by factorially combining these cues in three different synthetic speech continua.

Flege's work has concentrated on the differences between speakers as a function of their experience with final consonants. But the fact remains that, in the edited condition of experiment 2, all subjects made some use of vocalic cues. This may be explained by other work which suggests that the vocalic duration may be a universal voicing cue. One theoretical account that predicts universality of the vocalic duration effect in perception is the *perceptual contrast* hypothesis proposed by Kluender *et al.* (1988). Kluender *et al.* take the presence or absence of glottal vibration during closure to be the defining acoustic correlate of the voicing distinction in the context of VCV disyllables, but they consider closure duration to be an additional powerful cue. Vocalic duration cues voicing by enhancing the closure duration cue: Analogous to the case in vision, where a gray stimulus appears darker when viewed adjacent to a white stimulus than when viewed adjacent to a black one, in audition a long closure segment seems even longer when heard in the context of a short vocalic segment. Vocalic duration thus comes to cue voicing by enhancing the saliency of the closure interval duration: a short vocalic interval makes a long closure interval (i.e., the cue to − VOICE) seem even longer, and thereby favors perception of a final consonant as voiceless. Because the contrast hypothesis is based upon general principles of perception, it predicts that the capacity for vocalic duration to modify perception of closure duration (and thus that the ability to use vocalic duration as a voicing cue) should be universal. In support of this explanation of the vocalic duration effect, Kluender *et al.* report an experiment which manipulated "nonspeech analogues" of closure and vocalic duration. The parallel results obtained for perception of these analogues and speech were taken as evidence that the enhancement effect of vocalic duration is auditory in nature rather than being unique to the perception of speech signals.

The Kluender *et al.* study considered the vocalic duration contrast only within the context of VCV disyllables, but it seems that we could reasonably extend their theory to the CVC context, which is the concern of this paper. We might infer, from the ideas given in Kluender *et al.*, that non-native English speakers should be quite accurate in identifying English CVCs (e.g., "pod" or "pot") when the closure interval is marked by a release burst. Their theory could not have made a clear prediction as to whether non-native English speakers in experiment 2 should have been less able to use vocalic duration information to identify edited CVC tokens when closure interval and release bursts were removed and listeners had to rely on vocalic cues alone. For the present experiment, it predicts that all subjects will show some effect of vocalic duration on perception of final consonant voicing, but makes no prediction about possible group-related differences and any effect of experience. Also, their theory concerns only vocalic duration and makes no predictions about use of $F1$ offset frequency.

B. Method

1. Subjects

The subjects were the same 30 native English, Japanese, and Mandarin Chinese speakers who participated in experiments 1 and 2.

2. Stimuli

For experiment 3, three synthetic /pod/–/pot/ continua were synthesized using a version of the Klatt software package (Klatt, 1980). Parameter values were specified in 2-ms steps. Each continuum comprised nine stimuli that varied in vocalic duration; the three continua differed in $F1$ offset frequency (355, 455, or 555 Hz). Vocalic duration, including transitions, ranged from 100–260 ms in 20-ms steps. All stimuli began with a 50-ms interval of aspiration (AH = 50 dB). $F0$ began at 138 Hz and fell linearly to 95 Hz. For all stimuli, voicing source amplitude (AV) began at 0 dB and increased linearly to 57 dB over the first 50 ms (i.e., to the beginning of the steady state portion of the vocalic segment), and remained at this value until the final 50 ms where it began to decrease linearly to 40 dB at the end of the vocalic segment. None of the formants had an onset transition, and only $F1$ and $F2$ had offset transitions in the final 50 ms of the vocalic segment. See Table IV for other synthesis parameters.[3]

We pretested the stimuli in a pilot study which used subjects who had not participated in experiments 1 and 2. Native English speakers ($N = 7$) reliably perceived the stimuli as "pod" and "pot" and gave consistent identification functions. We then presented the stimuli to two non-native English speakers: one native Japanese listener and one native Mandarin Chinese listener who were somewhat more experienced than the Japanese and Mandarin subjects in experiments 1 and 2. Both of these speakers reported that, rather than perceiving a final consonant, they perceived only the initial consonant (the intended /p/) followed by shorter or longer duration versions of the vowel /a/. This being the case, we reasoned that the non-native English speaking subjects in our study would probably not be able to perform the task of identifying "pod" versus "pot."

We therefore sought a way to elicit a final consonant percept from our two non-native pilot subjects that would enable us to examine the effects of manipulating vocalic duration and $F1$ offset. This was accomplished by splicing a burst and closure interval (duration = 67 ms) from a natural token of /pod/ on to a zero-crossing of the last synthesized glottal pulse of each stimulus. After attenuation, the

TABLE IV. Formant values (in Hz) for the three synthetic "pod"–"pot" continua with low, medium, and high $F1$ offset frequencies. (IT = starting value of initial transition; SS = steady state frequency of vocalic segment; and FT = offset frequency of final transition. Formant bandwidths are in parentheses.)

Vocalic segment	$F1$	$F2$	$F3$
IT	675	1160	2460
	(60)	(90)	(150)
SS	675	1160	2460
	(60)	(90)	(150)
FT (low continuum)	355	1425	2460
	(60)	(90)	(150)
FT (medium continuum)	455	1425	2460
	(60)	(90)	(150)
FT (high continuum)	555	1425	2460
	(60)	(90)	(150)

amplitude of the burst was 12.8 mV. The resultant burst was barely audible relative to the vocalic segment (which was 363.7 mV), and not readily identifiable as either a /d/ or /t/ burst. These new stimuli were presented to the same Japanese and Mandarin subjects who had first heard the unaltered stimuli in the pilot study, as well as a new set of native English speakers ($N = 10$). Both the native and non-native pilot subjects now reported hearing a final consonant in the modified synthetic stimuli and could perform the task. The modified stimuli were therefore used in experiment 3. The stimuli were stored on the hard disk of an IBM AT computer, and low-pass filtered at 4.9 kHz prior to presentation.

3. Procedure

Experiment 3 followed experiment 2, and the subjects' task was the same: to identify each stimulus as either "pod" or "pot" by pressing a button marked "pod" or "pot" (and to guess if uncertain). The 27 tokens (nine vocalic durations by three $F1$ offset frequencies) were randomized and presented without replacement ten times each in randomized blocks using the Identify program (Shannon *et al.*, 1988) via TDH-39 earphones. The ISI and the "ready" signal were the same as in experiment 2.

C. Results and discussion

For each subject in each condition, the percentage of tokens identified as voiced ("pod") was recorded. Figure 2 shows the mean percentage of tokens identified as voiced, as a function of vocalic duration for each of the three different continua (three different $F1$ offset frequencies), separately for each of the three groups. The data were transformed using the arcsine function (see footnote 2) and analyzed with a group (native English, Japanese, or Mandarin listener) by $F1$ offset frequency (355, 455, or 555 Hz) by vocalic duration (nine different durations) repeated-measures MANOVA. The main effects of group [$F(2,27) = 6.47, p \ll 0.01$], $F1$ offset frequency [$F(2,26) = 37.7, p \ll 0.01$], and vocalic duration [$F(8,20) = 15.2, p \ll 0.01$] were all significant. The group by vocalic duration interaction [$F(16,42) = 2.4$, $p < 0.05$] was significant, but the group by $F1$ offset frequency interaction [$F(4,54) = 2.4, p = 0.063$] narrowly missed reaching significance. That is, sensitivity to the vocalic duration variable differed across subject group to a significant extent, but sensitivity to the $F1$ offset frequency variable did not. The fact that the group by $F1$ offset frequency interaction is nonsignificant may not be surprising, given that $F1$ offset, as compared to vocalic duration, may be a weaker voicing cue. The vocalic duration by $F1$ offset frequency interaction was nonsignificant. The three-way interaction of group by $F1$ offset frequency by vocalic duration was nonsignificant.

To assess between-groups differences in level of sensitivity to the vocalic duration cue, we collapsed the data over the $F1$ offset frequency variable and fit the resultant psychometric function for each subject with a normal ogive curve using probit analysis (see Finney, 1971). We estimated the slope of each subject's curve using seven of the nine levels of stimulus duration (end points of the continuum were excluded from analysis). We then submitted the estimated slope for each subject to a one-way ANOVA using native language as a grouping factor. The main effect of native language group [$F(2,27) = 11.6, p \ll 0.001$] was significant. *Post hoc* Newman–Keuls tests ($p = 0.05$) suggest that the mean slope was significantly greater for subjects in the native English group ($M = 13.0$, s.d. $= 5.7$; range [6.5, 24.7]) than for subjects in either the Japanese group ($M = 6.2$, s.d. $= 5.3$; range [-4.4, 13.6]) or the Mandarin Chinese group ($M = 3.3$, s.d. $= 2.2$; range [0.3, 7.6]). The mean slope was not significantly greater for subjects in the Japanese group than for subjects in the Mandarin group. The difference between the Japanese and Mandarin Chinese group would reach significance were we to exclude one of our Japanese subjects (subject KK). KK identified tokens with long vocalic segments as "pot" (instead of "pod") and tokens with short vocalic segments as "pod" (instead of "pot"), and thus generated a function similar in magnitude, but with a slope opposite in sign, to her groupmates. We did not surmise that KK had accidentally responsed "pot" when she meant "pod" because her data showed that, consistent with all of our other subjects, low $F1$ offset frequencies resulted in more voiced percepts, and high $F1$ offset frequencies resulted in more voiceless percepts. Because we only derived one function for each subject (and thus only ten functions for each group under study), KK's weight in the analysis was considerable, and thus added significantly to the variance within the Japanese group. Given this fact, it seems reasonable to allow that, compared to the Mandarin speakers, there is at least a tendency toward more sensitivity on the part of the native speakers of Japanese to the vocalic duration cue.

One might at this point raise the possibility that performance differences between the Japanese and Mandarin Chinese subjects did not result from differences in perceptual use of vocalic cues so much as from different response strategies. For example, it could be argued that the Japanese subjects were not categorizing the tokens according to the perceived final consonant, but were instead categorizing them according to the number of mora conveyed by the vocalic segment (see Flege, 1984, and Flege and Hillenbrand, 1986, for a similar viewpoint). That is, they may have re-

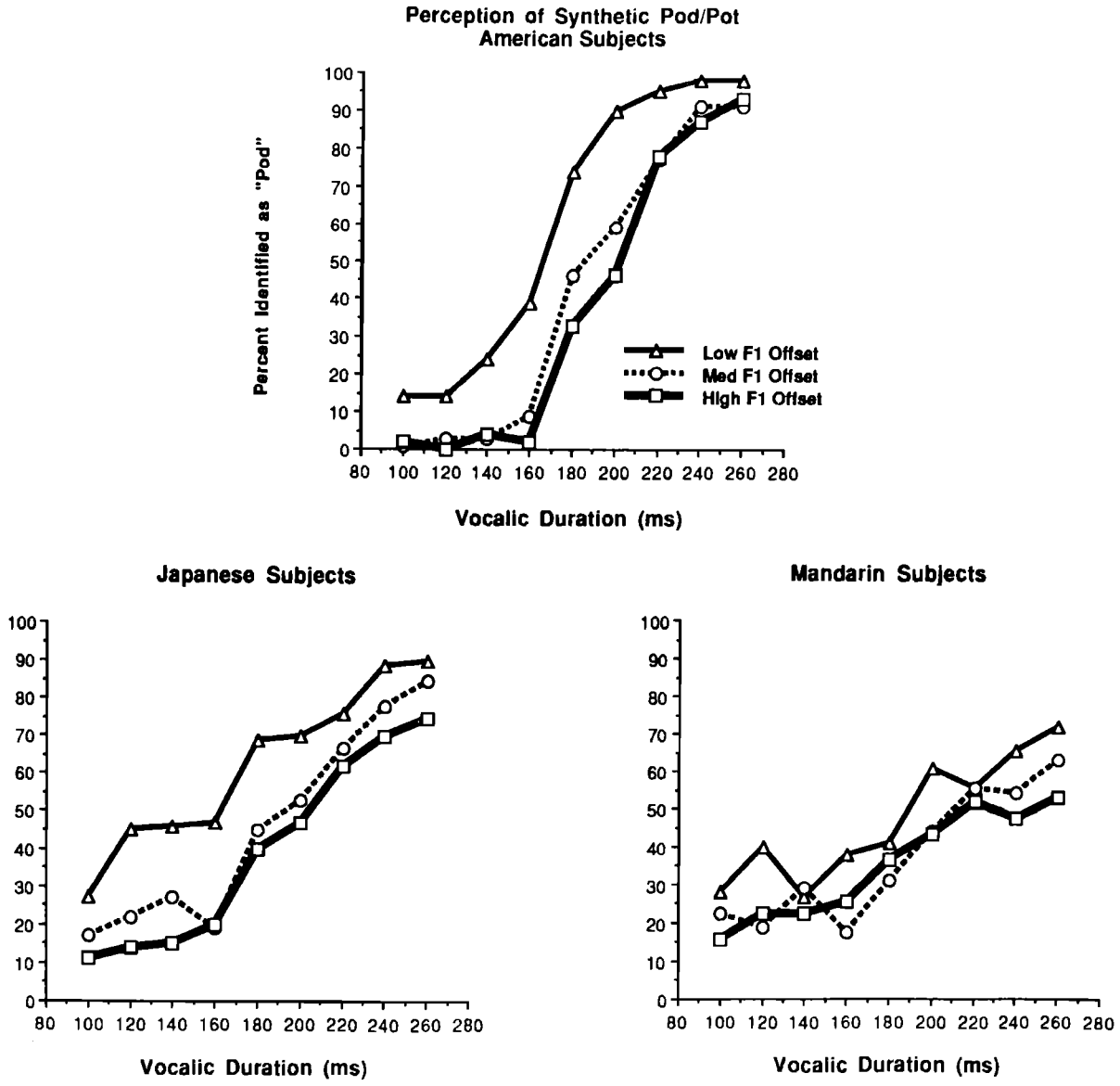


FIG. 2. Mean percentage of synthetic "pod"–"pot" tokens identified as "pod" as a function of subjects' native language background, vocalic duration, and $F1$ offset frequency. Vocalic duration ranged from 100 to 260 ms in 20-ms steps. Low, medium, and high $F1$ offset frequencies were, 355, 455, and 555 Hz, respectively.

sponded "pod" when the vocalic segment was two mora long, and "pot" when it was one mora long. Such a strategy would likely lead to the ogival functions we obtained for the Japanese subjects and perhaps for subjects in the other language groups, as well.

However, if we attribute the effects of duration to a mora-based strategy, we should also raise the question of why low $F1$ offset frequencies favored "two mora" responses, and high $F1$ offset frequencies favored "one mora" responses. It is possible that the $F1$ offset frequency of a vocalic segment influences its perceived length. If so, $F1$ may influence both perceived voicing and the perceived number of mora by modifying the perception of vocalic duration. We are currently examining these possibilities.

In the discussion for experiment 2, we raised the possibility that the performance difference observed between the Japanese group and the Mandarin Chinese group had resulted from differential sensitivity to certain vocalic cues. The results of experiment 3 now suggest that the Japanese and Mandrin subjects show differences in sensitivity to vocalic duration. Our extension of the Kluender *et al.* (1988) hypothesis to CVC syllables was consistent with the non-native English speakers being quite accurate in identifying natural, unedited pod/pot tokens in experiment 2. This experiment now suggests that, when closure and burst are present, subjects may make use of vocalic duration, but to different extents and as a function of native language background. To account for the different degree of sensitivity we observed between the native speakers of English and the Japanese and Mandarin subjects, a perceptual learning component must be included in our extension of the Kluender *et al.* (1988) perceptual contrast hypothesis.

IV. GENERAL DISCUSSION

In this study, we have seen how, for both production and perception of English CVC syllables, the extent to which subjects make use of vocalic cues to final consonant voicing appears to be a function of their native language experience. To summarize, for native speakers of English, Japanese, and Mandarin Chinese, vocalic duration is always longer before voiced consonants. Likewise, subjects in all three groups showed some sensitivity to vocalic duration as a perceptual voicing cue. However, the degree to which these speakers use vocalic duration in both production and perception varies systematically. Each experiment implies that, for both production and perception, the use of vocalic duration is greatest for native English speakers, least for native Mandarin Chinese speakers, and native Japanese speakers fall somewhere in between. We further observe that the $F1$ offset frequency values of CVCs produced by subjects in all three groups is lower before voiced consonants than voiceless consonants. Here, the size of the effect for both non-native English groups approached that of the native English subjects, and subjects did not exhibit a significant group related difference in degree of sensitivity to $F1$ offset as a perceptual voicing cue. Let us now discuss each cue, in turn, and the impact of each language-specific factor.

A. Vocalic duration

As we noted in the discussion of experiment 1, measurements of our subjects' productions of "pod" and "pot" are consistent with Chen's (1970) observation that, although there appears to be a universal tendency to employ a vocalic duration contrast, the magnitude of the contrast varies across languages. In Chen's study, failure to assess the contrast within the same consonantal context left open the question of whether syllabic context differences were responsible for differences between speakers of various languages. Experiment 1 now suggests that the magnitude of the contrast can vary according to native language group when subjects in each group attempt to produce the same CVC tokens.

An analogous conclusion can be drawn from experiment 3. Holding syllabic context constant across groups, perceptual sensitivity to vocalic duration is present for all subjects, but varies according to native language background. Our finding that each subject group demonstrated some sensitivity to vocalic duration is consistent with the possibility that, as implied in Kluender *et al.* (1988), vocalic duration is a universal perceptual cue. However, the extent of sensitivity appears to be influenced by language-specific factors.

Lack of experience with final stop consonants may help to explain the fact that the non-native English speakers produce smaller durational contrasts than their native English counterparts. Neither Japanese nor Mandarin allows CVC syllables ending in stops, and this could explain the tendency for native speakers of each of these languages to make less extensive use of vocalic duration than native speakers of English. As we noted in the Introduction, Flege and his colleagues (Flege, 1988, 1989; Flege and Wang, 1989) have observed reduced utilization of vocalic duration by Mandarin Chinese speakers as compared to speakers of other Chinese dialects. They suggest that lack of experience with final consonants is the basis of cross-linguistic differences in the use of vocalic cues. Flege's hypothesis may apply to our observation that Japanese and Mandarin speakers were less able to use vocalic duration than native speakers of English. If this is correct, a study of CVCV disyllables should yield no differences between them and native speakers of English. We are currently pursuing this possibility.

A focus on experience with final stop consonants does not really apply to observed differences between Japanese and Mandarin subjects. We have found that Japanese speakers produced a significantly greater durational contrast than the Mandarin speakers, were better able to identify CVCs, and were more sensitive to vocalic duration as a final consonant voicing cue. Yet the Mandarin speakers were, if anything, more experienced with English CVCs due to their longer residency in the US (see Table I). One way to further test the role of experience with English would be to replicate our study comparing very experienced (10 years residency in the US) native speakers of Japanese and Mandarin Chinese with inexperienced (2 months residency in the US) speakers of each language. If we are to explain the greater sensitivity of the Japanese subjects, at last one additional native language property must be considered, and native language experience with phonemic vowel length would seem to be a likely candidate. As we noted in the Introduction, we chose to compare Japanese and Mandarin Chinese speakers because Japanese, unlike Mandarin, includes a distinction between phonemically long and short vowels. This distinction is realized in the acoustic signal as a vocalic duration contrast.

In the Introduction, we had speculated, on the basis of Flege and Port's (1981) work on Arabic, that native experience with long versus short vowels might influence use of vocalic duration as a cue to final consonant voicing. However, where Flege and Port found that Arabic speakers seem to have almost no productive use of this cue, we have found that Japanese speakers do make use of this cue. Our findings would seem to imply that, for Japanese subjects, familiarity with phonemic use of long versus short vowels facilitates use of vocalic duration as a cue to final consonant voicing in English CVCs relative to Mandarin Chinese speakers, who lack native language familiarity with phonemic vocalic segment length distinctions. This interpretation would seem to predict that native Arabic subjects should share the Japanese advantage. One possible way to account for Flege and Port's Arabic findings is to consider the fact that, while Arabic employs a phonemic vocalic length contrast, it also permits final stop consonants, and these two factors could interact. Another is that experience with single versus geminate stop consonants is the basis of the Japanese advantage, as neither Arabic nor Mandarin permits geminates.

1. Relationship between productive and perceptual use of vocalic duration

As a final point in our discussion of use of vocalic duration, we should comment on the relationship between per-

ception and production. In our study, there is a concordance between use of vocalic duration in production and sensitivity to vocalic duration in perception: Native speakers of English show the strongest effect; Mandarin Chinese speakers show the weakest effect; and Japanese speakers fall in between. This raises certain questions about the relationship between productive and perceptual mastery of vocalic duration as a voicing cue. Some results in the literature imply that perceptual mastery may precede productive mastery (compare Flege and Port, 1981 with Flege, 1984), yet others have reported that speakers learning a second language may learn to produce a given contrast before they reliably perceive the same contrast (Sheldon and Strange, 1982; Gass, 1984). It may be generally true that speakers only produce the vocalic duration cues to the extent that they can perceive them. Our evidence is consistent with this possibility, both insofar as non-native speakers who are less able to produce the appropriate vocalic duration contrast are also less able to perceive it, and insofar as the relationship between native English, Japanese, and Mandarin subjects is the same for both production and perception. Yet we cannot rule out the possibility that failure to produce the appropriate vocalic duration contrast may only reflect the non-native speakers' inability to control the timing of his or her articulatory gestures in a native-English-like fashion. The fact that the Japanese subjects produced a more native-English-like vocalic duration contrast than the Mandarin subjects is consistent with the possibility that native experience with a phonemic vowel length distinction facilitated their ability to produce a more appropriate vocalic duration contrast in English CVCs. Longitudinal studies might help to decide between these alternatives.

B. F1 offset frequency

Previous research on native speakers of English has shown that $F1$ offset frequency is higher before voiceless consonants than voiced consonants, and that relative high and relatively low $F1$ offset frequencies cue voiceless and voiced final consonants, respectively (Fischer and Ohde, 1990; Summers, 1988; Walsh and Parker, 1983; Wolf, 1978; Hillenbrand *et al.*, 1984). Our innovation is to examine the use of $F1$ offset frequency as a function of native language experience with final stop consonants and phonemic vowel length. In experiment 1, we have found that native speakers of Japanese and Mandarin Chinese, who are learning English as a second language, produce vocalic segments with lower $F1$ offset frequencies in their tokens of "pod" compared to their tokens of "pot." They did not differ significantly from each other, but showed significantly less of an effect than native speakers of English. Both their difference from native speakers of English and their apparent similarity to each other are consistent with Flege's (1988) hypothesis that non-native English speakers who lack native experience with final stops will have difficulty with use of vocalic cues. That the Japanese and Mandarin subjects did not differ significantly suggests that the Japanese subjects' native language experience with phonemically long and short vowels does not influence their use of $F1$ offset frequency, which is as we speculated in the Introduction.

Experiment 2 does not distinguish perceptual use of $F1$ offset from use of other cues, although it may identify some difficulties on the part of the Japanese and Mandarin subjects. The perceptual use of $F1$ offset frequency was a direct concern of experiment 3, which showed that $F1$ offset affected perceived voicing for subjects in all three groups. Here, the group by $F1$ offset interaction narrowly missed reaching significance, and there are several possible explanations. First, because $F1$ offset may be a relatively weak voicing cue even for native English speakers, small between-group sensitivity differences may not have been resolved by the analysis we chose to use. A second possibility is that $F1$ offset may be a universal cue. Finally, $F1$ offset may be a language-specific cue that is more easily learned by non-native speakers. More cross-linguistic research will be needed to answer this question, especially studies employing very inexperienced subjects. Additional research will also be necessary to determine whether there may be an asymmetry between use of $F1$ offset frequency in production as opposed to perception.

C. Summary

In conclusion, we have found that native speakers of English, Japanese, and Mandarin Chinese share a tendency to use vocalic duration as a final consonant voicing cue in both production and perception of the English CVCs "pod" and "pot." However, they systematically differ in the degree to which they use this cue. They also share a tendency to use $F1$ offset frequency as a voicing cue. In this case, systematic differences in production seem clearer than differences in perception. To explain differences in use of vocalic duration, we point to the two factors of experience with final consonants and experience with phonemic use of vowel duration. Lack of experience with final stop consonants may hinder use of vocalic duration; experience with long versus short vowels may mitigate this problem. When we examine the role of these same factors in use of $F1$ offset frequency, the results are less clear, perhaps because $F1$ offset frequency is a weaker cue to final consonant voicing. Lack of experience with final stop consonants seems to hinder the use of $F1$ offset in production, and experience with long versus short vowels may not matter. Perceptual use of $F1$ offset may be less affected by native language experience or relatively easy to learn. Further research is necessary to evaluate these possibilities, and to determine whether the present results generalize to other stop consonants in other vowel environments.

ACKNOWLEDGMENTS

We thank Naoyuki Takagi for use of his English language experience questionnaire as well as his help with the pilot testing of experiment 3. We also thank James E. Flege, Terrance Nearey, and two anonymous reviewers for comments on an earlier version of this paper. This work was supported by a Pacific Rim grant to the University of California, Irvine. Portions of this paper were presented at the 120th Meeting of the Acoustical Society of America in San Diego, CA [J. Acoust. Soc. Am. Suppl. 1 **88**, S80 (1990)].

[1] Measurements of these additional materials have not yet been made and will be reported in a future study.

[2] The arcsine transform was used in order to stabilize the variance (see, for example, Bishop *et al.*, 1975; Kirk, 1982).

[3] Some of the parameter values and the format for the table were borrowed from Krause (1982), who studied the effect of vocalic duration on voicing perception.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (**1975**). *Discrete Multivairate Analysis: Theory and Practice* (MIT, Cambridge, MA).

Chen, M. (**1970**). "Vowel length variation as a function of the voicing of the consonant environment," Phonetica **22**, 129–159.

Finney, D. J. (**1971**). *Probit Analysis* (Cambridge, U.P., Cambridge), 3rd ed.

Fischer, R. M., and Ohde, R. N. (**1990**). "Spectral and duration properties of front vowels as cues to final stop-consonant voicing," J. Acoust. Soc. Am. **88**, 1250–1259.

Flege, J. E. (**1984**). "The effect of linguistic experience on Arabs' perception of the English /s/ vs. /z/ contrast," Folia Linguist. **18**, 117–138.

Flege, J. E. (**1988**). "The development of skill in producing word-final English stops: Kinematic parameters," J. Acoust. Soc. Am. **84**, 1639–1652.

Flege, J. E. (**1989**). "Chinese subjects' perception of the word-final English /t/–/d/ contrast: Performance before and after training," J. Acoust. Soc. Am. **86**, 1684–1697.

Flege, J. E., and Wang, C. (**1989**). "Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/–/d/ contrast," J. Phonet. **17**, 299–315.

Flege, J. E., and Hillenbrand, J. (**1986**). "Differential use of temporal cues to the /s/–/z/ contrast by native and non-native speakers of English," J. Acoust. Soc. Am. **79**, 508–517.

Flege, J. E., and Port, R. (**1981**). "Cross-language phonetic interference: Arabic to English," Lang. Speech **24**, 125–146.

Gass, S. (**1984**). "Development of speech perception and speech production abilities in adult second language learners," Appl. Psycholinguist. **5**, 51–74.

Hillenbrand, J., Ingrisano, D. R., Smith, B. L., and Flege, J. E. (**1984**). "Perception of the voiced–voiceless contrast in syllable-final stops," J. Acoust. Soc. Am. **76**, 18–26.

Homma, Y. (**1981**). "Durational relationship between Japanese stops and vowels," J. Phon. **9**, 273–281.

House, A. S. (**1961**). "On vowel duration in English," J. Acoust. Soc. Am. **33**, 1174–1177.

Keating, P. A. (**1985**). "Universal phonetics and the organization of grammars," in *Phonetic Linguistics: Essays in honor of Peter Ladefoged*, edited by V. A. Fromkin (Academic, Orlando, FL), pp. 115–132.

Kirk, R. E. (**1982**). *Experimental Design* (Brooks/Cole, Belmont, CA).

Klatt, D. H. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.

Kluender, K. R., Diehl, L., and Wright, B. A. (**1988**). "Vowel-length differences before voiced and voiceless consonants: an auditory explanation," J. Phon. **16**, 153–169.

Krause, S. E. (**1982**). "Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults," J. Acoust. Soc. Am. **71**, 990–995.

Mack, M. (**1982**). "Voicing-dependent vowel duration in English and French: Monolingual and bilingual production," J. Acoust. Soc. Am. **71**, 173–178.

Raphael, L. J. (**1972**). "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English," J. Acoust. Soc. Am. **51**, 1296–1303.

Shannon, V., Palumbo, R. L., and Grandgenett, M. (**1988**). "IDENTIFY: A program for speech recognition, identification, and categorization experiments on PC compatible computers," J. Acoust. Soc. Am. Suppl. 1 **83**, S16.

Sheldon, A., and Strange, W. (**1982**). "The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception," Appl. Psycholinguist. **3**, 243–261.

Stevens, K. N., Keyser, S. J., and Kawasaki, H. (**1986**). "Toward a phonetic and phonological theory of redundant features," in Invariance and Variability in Speech Processes, edited by J. S. Perkell and D. H. Klatt (Erlbaum, Hillsdale, NJ), pp. 426–449.

Summers, W. V. (**1987**). "Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses," J. Acoust. Soc. Am. **82**, 847–863.

Summers, W. V. (**1988**). "*F*1 structure provides information for final-consonant voicing,"J. Acoust. Soc. Am. **84**, 485–492.

Walsh, T., and Parker, F. (**1983**). "Vowel length and vowel transition cues to [+ / − voice] in post-vocalic stops," J. Phon. **11**, 407–412.

Wolf, C. G. (**1978**). "Voicing cues in English final stops," J. Phon. **6**, 299–309.