

Use of vocalic cues to consonant voicing and native language background: The influence of experimental design

COURT S. CROWTHER and VIRGINIA MANN
University of California, Irvine, California

For native speakers of English and several other languages, preceding vocalic duration and $F1$ offset frequency are two of the cues that convey the stop consonant voicing distinction in word-final position. For speakers learning English as a second language, there are indications that use of vocalic duration, but not $F1$ offset frequency, may be hindered by a lack of experience with phonemic (i.e., lexical) vowel length (the "phonemic vowel length account": Crowther & Mann, 1992). In this study, native speakers of Arabic, a language that includes a phonemic vowel length distinction, were tested for their use of vocalic duration and $F1$ offset in production and perception of the English consonant-vowel-consonant forms *pod* and *pot*. The phonemic vowel length hypothesis predicts that Arabic speakers should use vocalic duration extensively in production and perception. On the contrary, Experiment 1 revealed that, consistent with Flege and Port's (1981) findings, they produced only slightly (but significantly) longer vocalic segments in their *pod* tokens. It further indicated that their productions showed a significant variation in $F1$ offset as a function of final stop voicing. Perceptual sensitivity to vocalic duration and $F1$ offset as voicing cues was tested in two experiments. In Experiment 2, we employed a factorial combination of these two cues and a finely spaced vocalic duration continuum. Arabic speakers did not appear to be very sensitive to vocalic duration, but they were about as sensitive as native English speakers to $F1$ offset frequency. In Experiment 3, we employed a one-dimensional continuum of more widely spaced stimuli that varied only vocalic duration. Arabic speakers showed native-English-like sensitivity to vocalic duration. An explanation based on the perceptual anchor theory of context coding (Braida et al., 1984; Macmillan, 1987; Macmillan, Braida, & Goldberg, 1987) and phoneme perception theory (Schouten & Van Hoesen, 1992) is offered to reconcile the apparently contradictory perceptual findings. The explanation does not attribute native-English-like voicing perception to the Arabic subjects. The findings in this study call for a modification of the phonemic vowel length hypothesis.

The duration and first formant ($F1$) offset frequency of vocalic segments have been widely studied as cues to final stop consonant voicing for native speakers of English. Relatively long vocalic segments that terminate with relatively low $F1$ offset frequencies cue voiced final stop consonants, whereas shorter segments with higher $F1$ offsets cue voiceless final stop consonants (Raphael, 1972; Summers, 1987, 1988; Walsh & Parker, 1983; Wolf, 1978).

Questions remain, however, about the use of these two cues by speakers of other languages. Most of the previous research on the cross-linguistic use of vocalic duration as a voicing cue has focused primarily on production rather than on perception, and $F1$ offset frequency has not been studied extensively for either production or perception (Crowther & Mann, 1992). In the present study, we extended some of our recent work by testing productive and perceptual use of these two cues by native speakers of Arabic.

Portions of this paper were presented at the meeting of the Acoustical Society of America in Salt Lake City in May 1992. The authors wish to thank James E. Flege and three anonymous reviewers for comments that helped to improve the quality of this paper. They also thank Pat Keating for helpful ideas regarding the connection between $F1$ offset and voicing articulation, and Duncan Luce for suggestions regarding Experiment 3. Some of this work was completed as part of the first author's PhD dissertation from the Department of Cognitive Sciences at the University of California, Irvine, under the guidance of the second author. Correspondence should be addressed to C. S. Crowther, Phonetics Lab, Department of Linguistics, University of California, Los Angeles, CA 90024-1543.

Until fairly recently, most cross-linguistic data supported Chen's (1970) claim that there is a universal tendency to lengthen vocalic segments before voiced stop consonants. However, Flege and Port (1981) did not find a significant vocalic duration contrast in Arabic for native speakers of Arabic, and Keating (1985) found no measurable vocalic duration contrast in Czech for native speakers of Czech. There has been little cross-linguistic consideration of vocalic duration as a voicing cue in perception, but recent work with native speakers of Mandarin Chinese (Crowther & Mann, 1992; Flege, 1988, 1989; Flege & Wang, 1989) and native speakers of Japanese

—Accepted by previous editor, Charles W. Eriksen

(Crowther & Mann, 1992) suggests that the use of vocalic duration as a voicing cue may differ according to the native language background of the listener.

Crowther and Mann (1992) found that native speakers of Mandarin and Japanese use vocalic duration as a voicing cue in production and perception of *pod* and *pot*, but do so less extensively than native speakers of English. Because neither Mandarin nor Japanese allows word-final stop consonants, this finding is consistent with Flege's syllable processing hypothesis (Flege, 1989; Flege & Wang, 1989), which states that native experience with word-final stops may be required for listeners to be able to use multiple acoustic cues to identify word-final stop consonants. Interestingly, it was found that the Japanese subjects used vocalic duration in perception (and perhaps production as well) to a greater extent than did the Mandarin subjects. To explain this difference, the phonemic vowel length hypothesis was formulated by appealing to the fact that Japanese, but not Mandarin, includes a phonemic vowel length distinction. It was argued that the Japanese speakers somehow are able to draw upon their native language experience with the phonemic (i.e., meaningful, or lexical) use of vocalic segment length in the relatively novel situation of learning to use vocalic duration as a cue to final stop consonant voicing in English. This hypothesis is independently supported by more recent work (Flege, 1993), which shows that native speakers of English are more accurate in imitating the duration of auditorily presented isolated vowels than are native speakers of languages that do not use vocalic duration phonemically or as a voicing cue.

In the present study, we complemented our previous research by turning to Arabic, a language that includes a phonemic vowel length distinction and word-final stops as well. Although vocalic duration is probably not used as a voicing cue in Arabic consonant-vowel-consonant (hereafter, CVC) syllables (Flege & Port, 1981), the phonemic vowel length account nevertheless predicts that native Arabic speakers should use vocalic duration as a voicing cue in English, because Arabic includes a phonemic vowel length distinction. Some support for the possibility that experience with vowel length in Arabic facilitates the use of vocalic duration in a second language comes from recent findings of Munro (1990). He reports that native speakers of Arabic, as opposed to native speakers of English, more readily learn to identify synthetic tokens of the front rounded French vowels /y/ and /ø/ (which occur neither in English nor in Arabic) that are distinguished, in part, by vocalic duration. That the native Arabic speakers are better able to learn to distinguish a novel vowel pair on the basis of duration suggests that they may readily learn to make voicing decisions on the basis of vocalic duration differences in English.

It is especially important to test our hypothesis by using native speakers of Arabic because of the possibility that any facilitation due to native experience with phonemic vowel length is reduced by native experience with final

stop consonants. In learning to use vocalic duration as a voicing cue in English, native Japanese speakers may differ somewhat from native Arabic speakers. Japanese speakers do not have native experience with vocalic duration as a vowel cue in the CVC environment, so they must learn to use it as a voicing cue within a novel syllabic environment. Thus, native Japanese speakers learning English must learn to use vocalic duration in a new way and in a new environment. Because native Arabic speakers do have experience with word-final stops, it is possible that they would encounter interference if they needed to ignore vocalic duration as a vowel cue in order to use it as a voicing cue.

The literature relevant to the phonemic vowel length hypothesis prediction is difficult to interpret. Contrary to the phonemic vowel length hypothesis, Flege and Port (1981) report that native Arabic speakers produce only a small vocalic duration contrast between English CVCs ending in voiced as opposed to voiceless stop consonants. An attempt was made to replicate this result in Experiment 1. Consistent with the hypothesis, Flege (1984) reports perception data which suggest that native Arabic speakers may be quite able to use vocalic duration as a voicing cue for distinguishing the English words *peas* and *peace*. An attempt was made to replicate this finding in Experiments 2 and 3. The lack of agreement between production and perception may be at odds with Crowther and Mann's (1992) speculation that subjects should be expected to produce skillfully a vocalic duration contrast only to the extent that they are perceptually aware of the contrast.¹ If the Arabic subjects in Flege's (1984) *peas-peace* experiment were perceptually sensitive to vocalic duration as a voicing cue, it is unclear why subjects drawn from the same population in Flege and Port's (1981) study failed to evidence much of a duration contrast in their CVC productions. Extensive experience may be required for one to develop a native-English-like manner of timing one's articulatory gestures, even after one is perceptually aware of the contrast (Crowther & Mann, 1992; but see Sheldon & Strange, 1982, for what might qualify as a counterexample), but the striking degree of the asymmetry between production (Flege & Port, 1981) and perception (Flege, 1984) among Arabic speakers challenges even this line of reasoning.

One explanation for the apparent asymmetry is that the native Arabic subjects in Flege's (1984) study may have been unable to perceive the stimuli as *peas-peace*, and consequently adopted a response strategy that enabled them to complete the task. Flege himself suggests that although the Arabic subjects were behaviorally indistinguishable from the native English control subjects, they may have applied a strategy of identifying tokens in which the vowel was perceived as a phonemically long Arabic vowel as *peas*, and identifying those perceived as phonemically short as *peace*. Flege's explanation may well be correct, but other explanations can account for the data equally well. Hereafter, explanations such as Flege's that

do not attribute native-English-like voicing perception to native Arabic listeners will be referred to as *response-strategic* hypotheses.

The speech perception literature indicates that listeners' behavior depends on the range of acoustic parameter values constituting the stimulus continuum (see, e.g., Brady & Darwin, 1978; Rosen, 1979). In one condition of Flege's (1984) experiment, subjects were presented with 10 repetitions of five different synthetic tokens that varied only in vocalic duration. Vocalic duration in his *peas-peace* stimuli ranged from 150 to 350 msec, with a 50-msec step size. Because there were only five different stimuli, the subjects may have been able to memorize the label they applied to each stimulus and then, upon subsequent presentations, label each stimulus consistently and according to their memory of each stimulus-label association (Massaro & Cohen, 1976). This line of reasoning is consistent with Flege's data (see Figure 4 in Flege, 1984) and implies that native Arabic and English subjects' behavior may be differentially affected by experimental design factors such as the number of stimuli in each condition, parameter value ranges, continuum step sizes, the number of factors varied, and so forth. In the present study, this possibility was tested in two different perception experiments. In the first perceptual experiment, we employed a design in which vocalic duration and *F1* offset frequency were combined factorially, using a small step size along the vocalic duration dimension and more stimuli. In the second perceptual experiment, we employed stimuli that varied only along the vocalic duration dimension and mimicked Flege's (1984) *peas-peace* stimuli in number, step size, and range. If Arabic subjects employ a response strategy that does not involve native-English-like voicing perception, their performance, compared with that of native English subjects, may be influenced differently by the preceding design factors.

One goal in the present study was to test the predictions of the phonemic vowel length account by using native Arabic speakers. Another was to compare production with perception of vocalic duration as a voicing cue in order to determine whether there is an asymmetry between production and perception for native Arabic speakers.

The second area of interest in this study concerned *F1* offset frequency. It has been known for some time that when a stop is articulated after a vowel, the offset frequency of the vowel's first formant is generally lower when the stop is voiced than when it is voiceless (see, e.g., Wolf, 1978). More research is needed in order to determine whether changes occurring in *F1* offset as a function of voicing are a necessary consequence of post-vocalic stop articulation, or whether they reflect a language-specific practice (see the General Discussion).

F1 offset has not been studied widely in speakers of languages other than English. Recently, Crowther and Mann (1992) tested the use of *F1* offset as a voicing cue in the English CVCs *pod* and *pot* by native speakers of Mandarin Chinese and Japanese. The Japanese and Mandarin

subjects' productions showed systematic variation in *F1* offset as a function of voicing, although the magnitude of the variation was larger for native English speakers. They also demonstrated perceptual sensitivity to *F1* offset as a voicing cue, and group-related differences in sensitivity were nonsignificant. Thus, native language experience with a phonemic vowel length distinction may not impinge on the use of *F1* offset frequency as a voicing cue in English. Consequently, Crowther and Mann proposed that *F1* offset frequency may be either a universal (i.e., language-independent) voicing cue, or a language-specific voicing cue that is easily learned, perhaps owing to high auditory salience and/or ease of production. It was thought that the testing of native speakers of Arabic in Experiments 1 and 2 for use of *F1* offset frequency in production and perception would help us gain a better understanding of the use of the *F1* offset cue and its contrast with vocalic duration.

EXPERIMENT 1 Production of *Pod* and *Pot*

Experiment 1 concerned the use of vocalic duration and *F1* offset frequency in production of the English CVCs *pod* and *pot*. Flege and Port (1981) reported that native Arabic speakers produce, at best, a small vocalic duration contrast in English CVCs differing in final consonant voicing. Although their subjects did not produce the CVCs *pod* and *pot*, the native Arabic subjects in the present study were expected to produce a relatively small vocalic duration contrast between the words *pod* and *pot*.

Method

Subjects. The subjects in this experiment were 10 adult native Arabic speakers (2 females and 8 males) who arrived in the US after age 18. They were from Egypt ($n = 6$), Saudi Arabia ($n = 2$), or Jordan ($n = 2$). The median time that they had spent in the US was 10 months (see Table 1 for the overall time spent in the US for each subject; other information in the table will be introduced later). None was fluent in any language other than Arabic and English. They ranged in age from 18 to 52 years ($M = 30$). Nine of the subjects were either graduate or undergraduate students, and

Table 1
Relevant Subject Attributes and Performance Indices

Subject	TUS	Δ VD	Slope
1	2	5.4	3.6
2	2	2.1	2.8
3	3	-5.3	-0.5
4	6	19.2	1.6
5	8	11.7	5.8
6	12	37.4	1.3
7	14	16.0	2.3
8	48	10.6	5.2
9	96	43.8	4.2
10	116	109.8	5.5

Note—TUS, total time (in months) spent in the US; Δ VD, mean vocalic duration in *pod* tokens minus mean in *pot* tokens. Slope is the slope of the ogive functions in Experiment 2.

1 was the spouse of a graduate student. Six of the subjects had begun learning English in elementary school, and 4 had begun in high school. One had a native-English-speaking teacher for more than 6 months before arriving in the US. Four of them had listened to English educational programs for more than 6 months before coming to the US. Subjects were paid for their participation. Their data will be compared with data from the native English control group ($N = 10$) reported by Crowther and Mann (1992), who employed a procedure that was identical to the one in the present experiment.

Procedure. Subjects produced six tokens of *pod* and six tokens of *pot* embedded within the carrier phrase "Say _____ Chuck." They were not instructed to release the final stop. Their utterances were recorded on a Tascam cassette recorder (7.5 ips) with a Sony clip-on microphone. Before acoustic analysis, the tokens were digitized at 10 kHz and lowpass filtered at 4.9 kHz. The measurement procedure was identical to that employed in Experiment 1 in Crowther and Mann (1992). The vocalic segment, measured directly from the waveform, was defined as beginning after the /p/ release burst at the onset visible voicing and ending at the onset of stop closure as indicated by a decrease in amplitude and complexity of the waveform. *F1* offset frequency was measured using the formant tracking algorithm (20-msec Hamming window) included in the Interactive Laboratory System software. As in the procedure employed by Hillenbrand, Ingrisano, Smith, and Flege (1984), *F1* offset was defined by a decrease in *F1* peak amplitude and "the presence of abrupt discontinuities with previously extracted formant frequencies" (p. 23).

Results and Discussion

Vocalic Duration. A plot of the data for the native Arabic speakers, together with the analogous data for the native English speakers reported in Crowther and Mann (1992), is shown in Figure 1. A paired-sample *t* test revealed that the native Arabic-speaking subjects produced

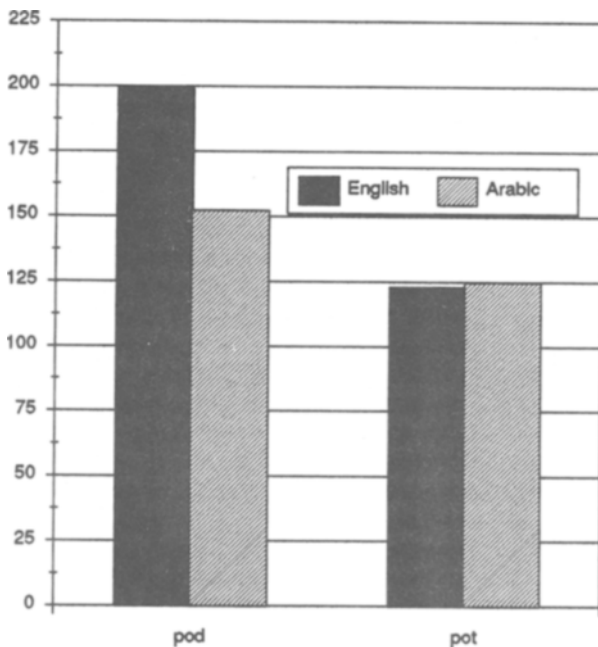


Figure 1. Mean vocalic duration (in milliseconds) of subjects' *pod* and *pot* tokens for native speakers of Arabic and native speakers of English.

significantly [$t(9) = 2.59, p < .05$] longer vocalic segments in their *pod* tokens ($M = 152.3$ msec, $SD = 85.7$) than in their *pot* tokens ($M = 124.9$ msec, $SD = 55.0$). The native English speakers also produced longer vocalic segments in *pod* tokens ($M = 198.9$ msec, $SD = 25.7$) than in their *pot* tokens ($M = 123.0$ msec, $SD = 27.3$). However, computing the vocalic duration contrast (mean vocalic duration in *pod* tokens minus mean vocalic duration in *pot* tokens) for each Arabic speaker and comparing the mean ($M = 27.4$ msec) with that for the native English speakers ($M = 75.9$ msec), we find that, consistent with Flege and Port (1981), the native English speakers produced a significantly greater durational contrast [$t(18) = 4.5, p < .01$]. The mean vocalic duration contrast for each native Arabic speaker is listed in Table 1. The difference in the magnitude of the vocalic duration contrast between native English and Arabic speakers is probably not due to group differences in speaking rate, because the vocalic duration ratio (mean vocalic duration before /d/ divided by mean vocalic duration before /t/) for the native English group (1.6) is substantially larger than that for the native Arabic group (1.2). Because the vocalic duration in *pot* tokens was about the same for both native language groups (just over about 120 msec), it appears that the between-group difference is due to a tendency on the part of the native English speakers to lengthen vocalic segments before /d/ more than the native Arabic speakers do.

With the possible exceptions of Subjects 9 and 10, there does not appear to be a strong tendency to produce a greater durational contrast with increasing experience in the US. Also, data from Subject 10 should be treated with caution, because the overall vocalic duration in his tokens greatly exceeded those of the typical native English subject and served to inflate the variance among the Arabic duration measurements. The vocalic duration in his *pod* ($M = 347.7$ msec) and *pot* ($M = 237.9$ msec) tokens may have been extended purposely. During debriefing, however, this subject reported that he was not aware of the English vocalic duration voicing effect, so interpreting his data is not straightforward.

***F1* offset frequency.** As in Crowther and Mann's (1992) study of native speakers of English, Mandarin, and Japanese, a paired-sample *t* test [$t(9) = 2.92, p = .017$] revealed that subjects in the Arabic group in this study produced significantly lower *F1* offset frequencies in their *pod* tokens ($M = 419$ Hz, $SD = 35.6$) than in their *pot* tokens ($M = 546$ Hz, $SD = 127.7$). However, the mean *F1* offset frequency contrast (mean *F1* offset in *pot* tokens minus mean *F1* offset in *pod* tokens) for the native Arabic speakers, although quite large ($M = 127$ Hz), is considerably smaller than that for the native English speakers ($M = 304$ Hz).

EXPERIMENT 2 Perception of *Pod* and *Pot*: A

In Experiment 2, we assessed subjects' perceptual sensitivity to vocalic duration and *F1* offset frequency by

using synthetic speech tokens intended to be perceived as *pod* or *pot*. It was noted in the introduction that subjects in Flege's (1984) perception experiment may have applied a strategy of first evaluating the vocalic segment duration and then making their voicing decision on the basis of that duration judgment. To discourage this strategy, we used more stimuli than Flege did (27, as opposed to 5 in his study), as well as a more finely grained vocalic duration continuum (20-msec step size, as opposed to 50 msec in his study). These measures were intended to make the possibility of memorizing the stimulus materials more difficult.

The inclusion of *F1* offset as an experimental parameter in and of itself may further help to discourage the response strategy noted earlier. If Arabic subjects are not sensitive to vocalic duration as a voicing cue, but are sensitive to *F1* offset as a voicing cue, they should be able to perform in a voicing perception task by making their voicing decision on the basis of *F1* offset information when vocalic duration and *F1* are varied factorially. Such a design will also increase the number of stimuli and thus increase the difficulty of implementing a response strategy that involves memorizing the stimulus materials. Testing sensitivity to both cues within a factorial design may thus serve to improve the design of the experiment (see Marsaro & Cohen, 1976, for a discussion of single factor vs. factorial designs in speech perception experiments).

Method

Subjects. The subjects who participated in Experiment 1 also participated in Experiment 2.

Stimuli and Procedure. The stimuli and procedure were identical to those of Experiment 3 in Crowther and Mann (1992). The stimuli consisted of three different nine-member *pod-pot* continua, each differing in *F1* offset frequency (355, 455, and 555 Hz). The members of each continuum were synthesized with a version of the Klatt synthesizer (Klatt, 1980), in cascade mode; they ranged in vocalic duration from 100 to 260 msec, in 20-msec steps. All stimuli began with a 5-msec burst appropriate for /p/. Fundamental frequency began at 138 Hz and fell linearly to 95 Hz by the end of the vocalic segment. A highly attenuated /d/ burst from a natural token of *pod* (after attenuation, the RMS amplitude of the burst was 12.8 mV and was barely audible relative to the vocalic segment, which had an RMS amplitude of 363.7 mV) and a 67-msec closure interval (energy in the closure interval was attenuated to zero) were spliced to the zero-crossing of the last glottal pulse of the vocalic segment of each token.

For testing, the 27 stimuli (nine levels of vocalic duration times three levels of *F1* offset frequency) were placed in random order in 10 different blocks. Before the subjects began the experiment proper, they listened to all 27 tokens presented in random order, to familiarize themselves with the task and the stimuli. In the test phase, subjects listened to the stimuli over earphones at a comfortable listening level and identified each token as either *pod* or *pot*. They responded by pressing a button marked *pod* or *pot*, and they were instructed to guess if uncertain. Each stimulus was presented 1.5 sec after the subject's response to the preceding stimulus, and presentation of the stimuli was controlled using the Identify software program (Shannon, Palumbo, & Grandgenett, 1988).

Results and Discussion

Vocalic duration. Figure 2 shows the mean percentage of *pod* identifications as a function of vocalic duration and *F1* offset frequency for subjects in the Arabic group and for the English subjects from Crowther and Mann (1992). To assess the subjects' sensitivity to vocalic duration, we employed the probit procedure used in Crowther and Mann. The data were collapsed over the *F1* offset variable for each native Arabic-speaking subject. The resultant psychometric function for each subject was then fit with a normal ogive by using probit analysis, and the mean slope for each subject was computed ($M = 3.2$; $SD = 2.1$; range, from -0.5 to 5.8); this is shown in Table 1. The slope for Subject 3 was negative and thus should be treated with some reservation. If "sensitivity to vocalic duration" is operationalized as the slope of each fitted ogive function, we should conclude that the native Arabic subjects are not very sensitive to vocalic duration, because the mean slope for subjects in their group is rather small. This slope value is significantly less than the value obtained for the native English group ($M = 13.0$; $SD = 5.7$; range, from 6.5 to 24.7) in Crowther and Mann, which suggests that the native English speakers were significantly more sensitive than the native Arabic speakers were to vocalic duration as a voicing cue for distinguishing *pod* from *pot*. The native Arabic subjects may be less sensitive than native Japanese speakers ($M = 6.2$), but about as sensitive as Mandarin speakers ($M = 3.3$) (see Crowther & Mann, 1992). This finding conflicts with predictions of the phonemic vowel length account.

***F1* offset frequency.** To assess the native Arabic subjects' sensitivity to *F1* offset frequency as a voicing cue, a paired-sample *t* test was used to compare each subject's responses at each level of the vocalic duration variable for the lowest versus highest level of the *F1* offset variable. The tests revealed that the *F1* variable significantly affected their voicing judgments [$t(89) = 9.1$, $p < .01$]. Collapsing the data over the vocalic duration variable, the mean percentages of *pod* identifications for native English speakers are 60.7, 42.2, and 38.3, respectively, for the low, medium, and high levels of the *F1* offset variable. For the native Arabic speakers, the mean percentages of *pod* identifications are 63.1, 47.4, and 41.3, respectively, for the low, medium, and high levels of the *F1* offset variable. To compare the influence of *F1* as a voicing cue for native Arabic speakers to its influence for native English speakers, we computed an *F1* impact score (the total number of voiced responses for the lowest level of the *F1* offset variable minus the total number of voiced responses for the highest level of *F1*) for each subject in the Arabic and English groups. An independent sample *t* test failed to reveal a significant difference between the *F1* impact score for native English ($M = 20.1$, $SD = 11.9$) and that for native Arabic ($M = 19.6$, $SD = 12.9$) speakers [$t(18) = 0.09$, $p > .9$].

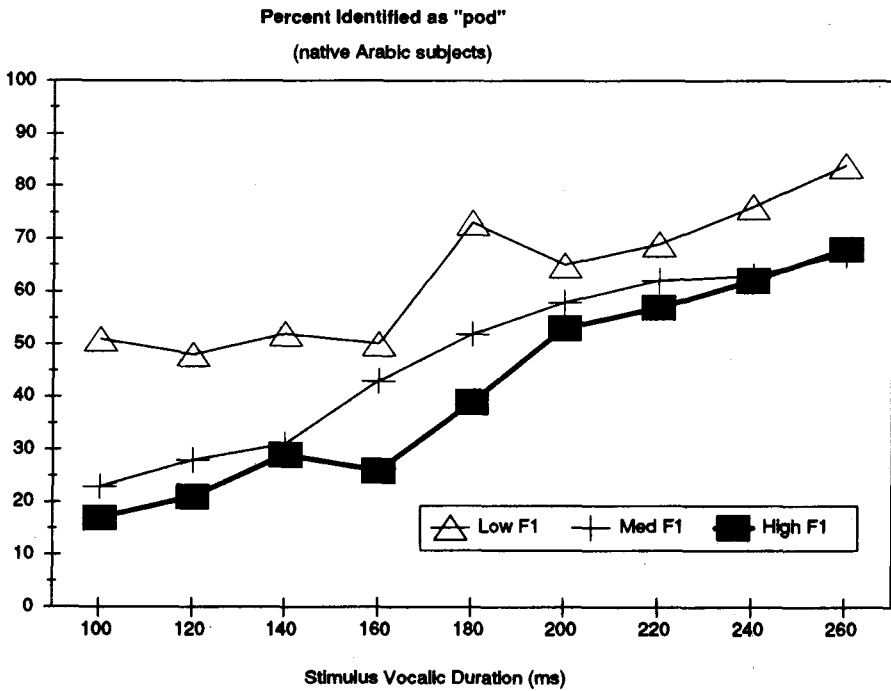
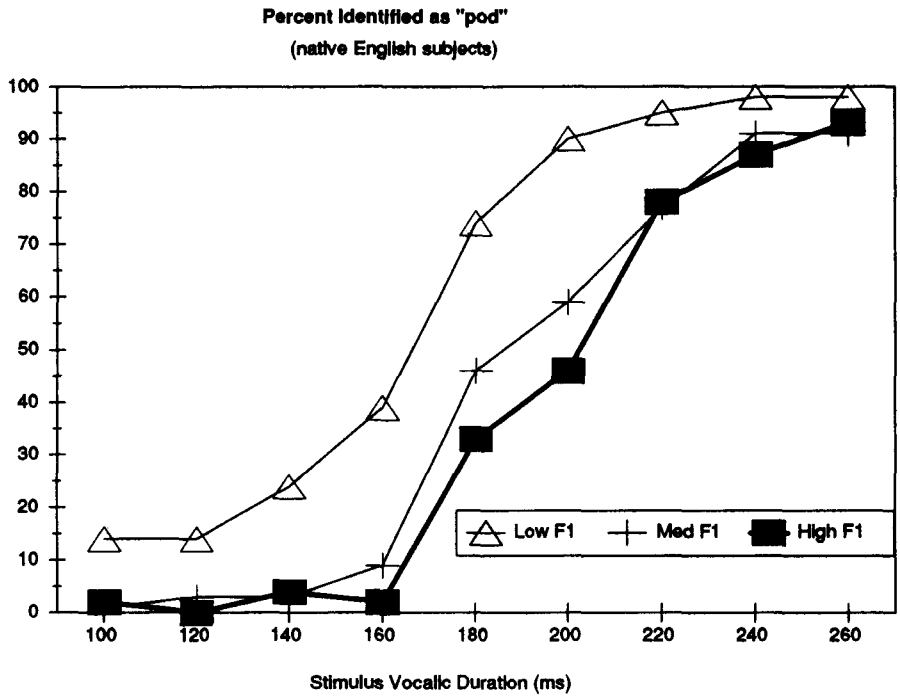


Figure 2. Mean percentage of synthetic *pod-pot* tokens from Experiment 2 identified as *pod* as a function of vocalic duration and *F1* offset frequency. Vocalic duration ranges from 100 to 260 msec in 20-msec steps. Low, medium, and high *F1* offset frequencies are 355, 455, and 555 Hz, respectively.

The native Arabic speakers are therefore sensitive to $F1$ offset frequency as a voicing cue, and their sensitivity may not differ significantly in magnitude from that of native English speakers.

EXPERIMENT 3 Perception of *Pod* and *Pot*: B

The results of Experiment 2 are in conflict with Flege's (1984) findings, in suggesting that native Arabic speakers are far less sensitive than native English speakers to vocalic duration as a voicing cue. It was proposed in the introduction that Flege's findings may not reflect the Arabic subjects' sensitivity to vocalic duration as a voicing cue, in part because the nature of his design may have allowed the Arabic subjects to apply a categorization strategy based on vocalic length *per se*, rather than on voicing perception. The results of Experiment 2 may be more representative of the Arabic speakers' sensitivity, because the nature of the stimulus materials (smaller step size, more stimuli, and factorial combination of two different voicing cues) may have made it more difficult for the subjects to apply any one of the response strategies outlined above. Alternatively, it may be the case that the vocalic duration cue operates differently for voicing in the fricatives (/s/-/z/) studied by Flege (1984) than for voicing in the alveolar stops (/d/-/t/). Sensitivity to vocalic duration as a voicing cue for fricatives may not necessarily imply sensitivity to vocalic duration as a voicing cue for stops.

One way to test the response strategy hypothesis is to attempt to replicate Flege's experiment by modifying the

pod-pot stimuli from Experiment 2 to have the same step size that he used (50 msec), and only five stimuli that vary in vocalic duration. If the native Arabic speakers generate native-English-like data when the stimuli are modified in this way, a response strategy explanation will be supported. This was the course taken in Experiment 3.

Method

Subjects. The subjects in this experiment include 7 adult native Arabic speakers who began living in the US after age 18. None had participated in Experiment 2. The median time that they had spent in the US was 8 months (range, 4-30 months). None was fluent in any language other than Arabic and English. All 7 subjects had begun learning English in elementary school. They were paid for their participation.

Seven native English speakers who had not participated in Experiment 3 served as a control group. They were given extra credit in an undergraduate social science course at the University of California, Irvine, as compensation for participating.

Stimuli and Procedure. To create the stimuli for Experiment 3, we used the same synthesis techniques as those described in Experiment 2, but with a few exceptions intended to make them mimic the temporal properties of Flege's (1984) stimuli. The level of the $F1$ offset frequency variable was fixed at 455 Hz (the intermediate level of the three levels used in Experiment 2), because this appears to be the most neutral of the three levels in terms of voicing. Levels of the vocalic duration variable were adjusted to range from 150 to 350 Hz, in 50-msec steps. Thus, these stimuli are analogous (in number, vocalic duration range, and step size) to Flege's *peas-peace* stimuli.

The procedure was the same as in Experiment 2. The subjects first listened to the five tokens played out in random order to familiarize themselves with the stimuli and the task. In the test phase, the subjects listened to the stimuli (10 times each, in random order) over earphones at a comfortable listening level and identified each token as either *pod* or *pot*. They were instructed to guess if uncertain.

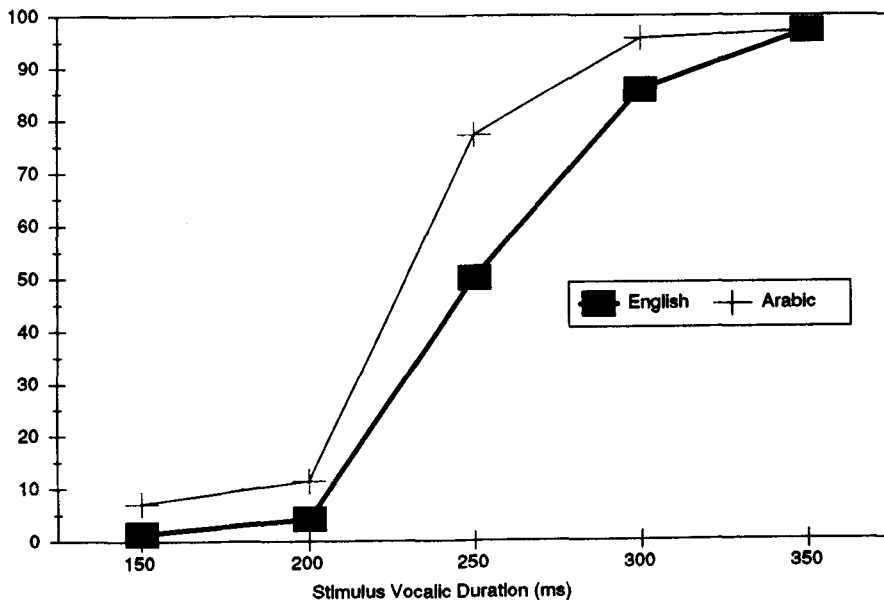


Figure 3. Mean percentage of synthetic *pod-pot* tokens from Experiment 3 identified as *pod* as a function of vocalic duration, which ranges from 150 to 350 msec, in 50-msec steps. $F1$ offset frequency is fixed at 455 Hz.

Results and Discussion

Figure 3 plots the mean percentage of *pod* identifications as a function of vocalic duration for subjects in the native Arabic and native English groups. Whereas the Arabic and English data from Experiment 2 appear to be quite different (see Figure 2), their data from Experiment 3 appear to be far more similar, particularly for the endpoint stimuli. In Experiment 3, the native Arabic subjects, like the native English subjects, almost unanimously identified the shorter stimuli as *pot* and the longer stimuli as *pod*, which could imply that they were sensitive to vocalic duration as a voicing cue. For Experiment 2, however, the slope of the Arabic subjects' identification function is shallow ($M = 3.2$, and see Figure 2) and they did not label the endpoint stimuli consistently, suggesting that they were not very sensitive to vocalic duration.

One interpretation of the apparently contradictory results of Experiments 2 and 3 is that the data from Flege's *peas-peace* experiment and the present Experiment 3 reflect a response strategy that does not require voicing perception for its implementation. Following this line of reasoning, the data from Experiment 2, in which we used a larger step size, more stimuli, and a factorial combination of two voicing cues, may more accurately reflect the Arabic subjects' sensitivity to vocalic duration as a voicing cue.

Another explanation for the discrepancy in the Arabic data between Experiments 2 and 3 is based on the fact that in Experiment 2 vocalic duration varied factorially with *F1* offset, but in Experiment 3 it varied alone. In Experiment 2, the subjects may have made their voicing judgments on the basis of *F1* offset frequency, but in Experiment 3 they had to rely on vocalic duration alone and thus may have focused more attention on vocalic duration. Intuitively, had this been the case, the *F1* offset variable should have influenced voicing judgments to a greater extent than it actually did. In any case, it seems that vocalic duration probably impacted voicing decisions in Experiment 2 more than did *F1* offset (see Figure 2), making this explanation implausible.

GENERAL DISCUSSION

This study shows that the vocalic duration in tokens produced by native speakers of Arabic varied to a small extent as a function of stop consonant voicing. The perceptual results are less clear: native Arabic speakers appeared to be quite sensitive to vocalic duration as a voicing cue when the vocalic duration continuum contained only five widely spaced members (Experiment 3), but showed much less sensitivity when vocalic duration and *F1* offset were varied factorially in three more finely spaced continua (Experiment 2). In contrast, they may be as sensitive as native English speakers to *F1* offset frequency as a voicing cue in perception, although they probably use it in production to a lesser extent than do their native English counterparts.

The Findings for Production: Theoretical Implications

In Experiment 1, it was found that the native Arabic subjects produced a vocalic duration contrast of about 27 msec and an *F1* offset frequency contrast of about 127 Hz. The sizes of these contrasts are smaller than those found with native English speakers, but they are certainly large enough to be discriminable under ideal listening conditions, such as those employed in typical speech perception experiments. Fujisaki, Nakamura, and Imoto (1975) report a just noticeable difference (JND) of about 10 msec for vocalic duration in isolated vowels, and Mermelstein (1978a) reports JNDs of about 60 Hz for the *F1* steady-state frequency of vowels embedded in CVC syllables.

The implications of the sizes of the vocalic duration and *F1* offset frequency contrast differences between native English and native Arabic speakers depends on the nature of the speech perception theory in question. For auditory contrast theories (see, e.g., Kluender, Diehl, & Wright, 1988; Parker, Diehl, & Kluender, 1986), which are based on general psychoacoustic principles, even small duration and frequency differences may influence speech perception to some extent, as long as they are noticeable under everyday listening conditions: "Experienced listeners make use of *all* potentially relevant cues for phonetic categories, provided [that] these cues are detectable" (Diehl & Kluender, 1987, p. 226). However, for prototype matching theories (e.g., Massaro & Oden, 1980; Oden & Massaro, 1978) and theories that explicitly relate perceptual effects of speech cues to the distributions of these cues in production (e.g., Nearey & Hogan, 1986), one should expect that statistically large differences in the production distributions of acoustic cues may be required if these cues are to be used in perception. For example, assume that the distribution of vocalic durations in /ad/ and /at/ syllables overlapped to a very great extent for speakers in some language community. If listeners make voicing judgments through reference to internalized representations of those distributions, there is a high probability that listeners would identify /ad/ syllables as /at/ syllables, and vice versa, because it would be difficult or impossible to determine the distribution from which the sample was drawn. Of course, in cases in which cue distributions overlap, these models could de-emphasize the importance of the cues by either ignoring them altogether or assigning them a relatively small weight in perception. The size of the acoustic parameter contrasts must therefore be interpreted with respect to the nature of the model in question. For auditory contrast models, the magnitude of the contrast may not be as important as it is for prototype matching models and for statistical decision models.

Vocalic Duration

The discrepancy between Experiments 2 and 3. The results of Experiment 2 suggest that native Arabic speakers are much less sensitive than native English speakers to vocalic duration as a voicing cue, but those

of Experiment 3 suggest that the two groups are quite similar in sensitivity. There are several plausible explanations for this discrepancy. First, the vocalic duration continuum ranged from 100 to 260 msec in Experiment 2, and from 150 to 350 msec in Experiment 3. As noted in the introduction, Flege (Flege, 1984; Flege & Hillenbrand, 1986) suggests that Arabic subjects in voicing perception experiments may not perceive the stimuli in a native-English-like manner, but instead may identify the stimuli on the basis of the duration of Arabic long/short vowels: stimuli perceived as containing long vowels are identified as voiced, and those perceived as containing short vowels, as voiceless. In terms of this scenario, the results of Experiment 2 would be attributed to an inability on the part of the Arabic subjects to identify the stimuli according to the Arabic long/short vowel distinction, because the endpoints of the vocalic duration continuum were not representative of Arabic long and short vowels. They were able to use the long/short vowel strategy in Experiment 3 because the endpoints may have corresponded better to Arabic long/short vowels.

There is a more general psychophysical explanation of the discrepancy that applies some of the findings and concepts of the perceptual anchor theory of context coding (Braidia et al., 1984; Macmillan, 1987; Macmillan, Braidia, & Goldberg, 1987; Macmillan, Goldberg, & Braidia, 1988). Although formulated originally for auditory intensity perception, the theory would seem to be applicable to other tasks. In terms of this theory, stimuli at the endpoints of the vocalic duration continuum, like those at the endpoints of an intensity continuum, serve as perceptual anchors, and subjects categorize each stimulus by measuring the perceived distance between it and each anchor. The particular stimuli employed in an experiment thus define a context, and subjects consider each stimulus according to its position in the overall stimulus context. The step size in Experiment 3 and in Flege's *peas-peace* experiment (50 msec) was considerably larger than that in Experiment 2 (20 msec). Comparisons between a stimulus and a perceptual anchor are made with less variance when the step size is large (Braidia et al., 1984); hence, category assignments were made with less variance in Experiment 3 than in Experiment 2.

The relatively simple nature of the stimulus continua in Experiment 3 and Flege (1984) would seem to make the preceding explanation especially plausible. The just noticeable difference for isolated vowels is only about 10 msec (Fujisaki et al., 1975), so a 50-msec step size would probably make the vocalic duration differences between stimuli quite salient. Arabic subjects, being skilled at exploiting vocalic duration differences to identify vowels (Munro, 1990), and faced with perceptually salient vocalic duration differences, may easily have adopted a strategy of selectively attending to vocalic duration (see Schouten & Van Hessen, 1992) and categorizing the stimuli with the two longest vocalic segments as *peas*, and those with the two shortest segments as *peace*. They may have categorized the stimulus with intermediate vocalic

segment length as *peace* or *peas* less consistently, because it was far from both perceptual anchors (Macmillan et al., 1988). Intuitively, a strategy of comparing 5 different stimuli with perceptual anchors (Experiment 3; and Flege, 1984) should be easier than comparing 27 different stimuli with perceptual anchors (Experiment 2), so it was probably easier for subjects to maintain a consistent response strategy in Experiment 3 than in Experiment 2.

The perceptual anchor explanation accounts well for the discrepancies in the Arabic subjects' data between Experiments 2 and 3, but one potential problem is that if it is applied to the native Arabic data, it should also be applied to the native English data. The native Arabic data evidenced ceiling effects for endpoint stimuli only in Flege (1984) and Experiment 3, but the native English data evidenced ceiling effects in both Experiments 2 and 3. The theory does not predict that experimental design factors should affect each subject group differently. However, if "context coding of consonants is with reference to well-remembered prototypes" (Macmillan et al., 1987, p. 35), it may be the case that the native Arabic speakers do not possess prototypes that are as well defined as those of native English speakers (see, e.g., Flege & Hillenbrand, 1986), which might explain why the subject groups were affected differently by experimental design factors.

A recent extension of the perceptual anchor theory may provide a more detailed framework for explaining performance differences between Arabic and English subjects in Experiments 2 and 3. Schouten and Van Hessen extended the perceptual anchor theory by adding, among other things, a long-term phoneme memory for stop consonants (Schouten & Van Hessen, 1992; Van Hessen & Schouten, 1992). According to their phoneme perception theory, listeners possess long-term-memory representations of the phonemes that occur in their native language. The representations are language specific, in that they depend critically on the native language experience of the listener. Given this theory, the discrepancy in the native Arabic listeners' results between Experiments 2 and 3 is not surprising. If their phoneme representations (or prototypes) are different from those possessed by native English listeners, or if they lack relevant representations altogether, their response behavior should be expected to be influenced more by the nature of the stimulus materials (e.g., step size, range, number of stimuli). Although the procedures were identical for subjects in both groups, the tasks in Experiments 2 and 3 may have been psychologically different for the native English and Arabic listeners, in that the native English listeners may have categorized the stimuli according to long-term-memory representations (developed over the course of English acquisition) of /d/ and /t/, whereas the native Arabic listeners, lacking a native English background, may have been compelled to categorize the stimuli according to perceptual anchors that were "induced" during the experiments. In terms of phoneme perception theory, then, native Arabic listeners may have relied more on the overall stimulus context than did their native English counterparts,

who presumably classify stop consonants according to phoneme memory (if the stimuli are good exemplars of real speech). Consequently, the performance of subjects in both groups was affected differently by experimental design factors.

It should be noted that this explanation is consistent with prototype matching models (see, e.g., Massaro & Oden, 1980; Oden & Massaro, 1978), and with models that explicitly relate perception behavior to the distribution of segmental properties (e.g., voice onset time, vocalic duration) of speech sounds within a language community (see, e.g., Nearey & Hogan, 1986). Flege's account (Flege, 1984; Flege & Hillenbrand, 1986) and the perceptual anchor account cannot be distinguished on the basis of findings in this study, although they are empirically distinguishable. Flege's account suggests that native Arabic subjects make their voicing decisions on the basis of Arabic long/short vowel perception. This would seem to predict that speakers of languages that do not include a vowel length distinction and do not use vocalic duration as a voicing cue (such as native Mandarin speakers) should not be expected to show the performance differences that native Arabic speakers showed when tested in Experiments 2 and 3 above. Another empirical test would be to replace the vocalic portion of the stimuli in Experiment 3 with noise, and test a new group of Arabic subjects. Finding that their psychometric functions are shallower than those in Experiment 3 would support Flege's hypothesis.

Failure of the predictions for Arabic use of vocalic duration. The phonemic vowel length hypothesis (Crowther & Mann, 1992) predicted that native Arabic subjects should use vocalic duration extensively, because Arabic includes a phonemic vowel length distinction and word-final stop consonants (Flege, 1988, 1989; Flege & Wang, 1989). In Experiment 1, subjects produced a small (but significant) vocalic duration contrast between *pod* and *pot* tokens, replicating Flege and Port (1981). Also contrary to the prediction, subjects in Experiment 2 showed far less sensitivity than native English speakers to vocalic duration as a voicing cue.

Native language experience with phonemic vowel length seems to give Japanese speakers an advantage for use of vocalic duration as a voicing cue in English (Crowther & Mann, 1992), but does not seem to do so for the Arabic subjects. In fact, native Arabic speakers are about as poor as native Mandarin speakers in their use of vocalic duration, even though Mandarin does not include a phonemic vowel length distinction or word-final stops. The phonemic vowel length account may be salvaged by constraining it to cases in which speakers lack experience with final stop consonants. The CVC (final stop) syllable is a novel syllable for the Japanese subjects, and thus, when learning English, they must learn to use vocalic duration in a novel way (as a voicing cue) within a novel syllabic environment (the CVC syllable). For native Arabic speakers, however, the CVC (final stop) syllable is not a novel syllable: Arabic CVC syllables contrast final voiced and

voiceless stops, just as in English. However, in Arabic the voicing contrast is not conveyed by vocalic duration. Flege and Port (1981) reported no significant vocalic duration contrast for two different Arabic minimal CVC pairs differing in final voicing. Thus the Arabic subjects may be accustomed to using vocalic duration within the CVC context only as a cue to vowel identity, and to using other cues to consonant voicing, such as the presence or absence of periodic vibration during closure (see Flege & Port, 1981). Consequently, for native Arabic speakers learning English, the process of learning to use vocalic duration in a native-English-like manner (i.e., as a voicing cue in the CVC context) may be inhibited as a result of the more familiar use of vocalic duration within the CVC context (i.e., as a cue to vowel identity).

Given the preceding empirical findings, the phonemic vowel length hypothesis should be constrained. Native experience with vocalic duration contrast as a cue to vowel identity may either facilitate or inhibit the process of learning to use the durational contrast in a novel way. If the novel application for the durational contrast occurs within an unfamiliar syllabic context, the learning process will be facilitated. If it occurs within a familiar syllabic context, the learning process may not be facilitated.

One potential difficulty for this modification of the phonemic vowel length account follows from studies showing that native English speakers are able to use vocalic duration in two different ways. Not only can they use it as a voicing cue to distinguish minimal pair CVCs such as *bad* and *bat*, but also as a vowel cue to distinguish minimal pair CVCs such as *bat* and *bet* (Crowther, 1994; Mermelstein, 1978b; Whalen, 1989). It is therefore unclear why the native Arabic speakers are able to use vocalic duration in one way (as a vowel cue), but have difficulty learning to use it in a second way (as a voicing cue) within the familiar CVC context. To attain mastery of vocalic duration as a cue to both vowel identity and voicing, perhaps native exposure is required.

Correspondence between productive and perceptual use of vocalic duration. The apparent asymmetry between the use of vocalic duration in production and perception for native Arabic speakers (compare Flege & Port, 1981, with Flege, 1984) was inconsistent with observations in Crowther and Mann (1992). The present study suggests that native Arabic speakers, like native speakers of Mandarin and Japanese, may use vocalic duration in production to the extent that they are sensitive to vocalic duration in perception. The finding of parallels between production and perception should not be surprising when one is considering speakers who learn English as a second language in a "naturalistic" manner. In general, it is not clear how a speaker could attain productive mastery of a distinction (e.g., the English vocalic duration/voicing relationship) without perceptual awareness of the distinction (but see Gass, 1984; Sheldon & Strange, 1982). Of course, in the case of the association between vocalic duration and voicing, some have argued (e.g., Chen, 1970) that vocalic segments are necessarily longer before

voiced stops because of particularities of human vocal tract anatomy. If the vocalic duration contrast is a by-product of postvocalic stop voicing production, perhaps one may produce a vocalic duration contrast without perceptual awareness of the contrast (but see Kluender et al., 1988, for a critical review of several "articulatory necessity" explanations).

F1 Offset Frequency

The native Arabic subjects seemed to use *F1* offset frequency extensively and systematically as a cue to final consonant voicing in both production and perception. Crowther and Mann (1992) report substantial use of *F1* offset frequency by native speakers of English, Japanese, and Mandarin Chinese, which suggests that the native language factors in question in this study, inclusion of a phonemic vowel length distinction and presence of word-final stop consonants, may be neutral to use of *F1* offset as a voicing cue when one is learning English as a second language.

These findings, together with those in Crowther and Mann (1992), are consistent with the possibility that *F1* offset may be learned easily, perhaps because of high auditory salience and ease in production. Some have suggested that glottal pulsing during closure of C_2 in a C_1VC_2 syllable (where C_2 is a stop consonant) is the essential cue to voicing (Kluender et al., 1988; Stevens, Keyser, & Kawasaki, 1986). If this was the case, listeners might interpret a low *F1* offset transition as vocal cord vibration during closure, which would explain why low *F1* offsets favor voiced percepts.

For production, initiation of postvocalic closure voicing for a voiced stop may entail articulatory gestures (supraglottal cavity expansion; tongue-raising gesture) that serve to lower *F1* offset frequency, in which case finding a relationship between voicing and *F1* offset across many different language communities should not be surprising. Kent and Moll (1969) showed that the supraglottal cavity expands actively during production of voiced stops (see also Westbury, 1983), so perhaps the articulatory adjustments required to effect supraglottal cavity expansion for voiced stops result in a lower terminating value for *F1*. Hillenbrand et al. (1984) refer to the "timing of a phonation-terminating gesture relative to the achievement of articulatory closure" as "voice offset time" (p. 18). Because voice offset time is earlier (perhaps before, or coincidental with, the initiation of closure) for voiceless than for voiced stops, *F1* offset frequency is generally higher for voiceless stops (Hillenbrand et al., 1984). On the other hand, *F1* tends to decrease during articulatory narrowing (Fant, 1962) and is associated with tongue height during vowel articulation—a vowel produced with the tongue at a higher position in the mouth, and therefore with a narrower aperture, has a lower *F1* steady-state frequency (Ladefoged, 1982). Consequently, one might speculate that postvocalic voiced stops are produced with the tongue in a higher position than it is for homorganic voiceless stops. Summers (1987) provides a somewhat in-

direct test of this hypothesis, in that he found a statistically significant relationship between jaw height at vowel offset and final stop consonant voicing for only 1 of 3 native English subjects. However, failure to find a robust relationship between jaw position and *F1* offset frequency does not disconfirm the hypothesis that tongue height varies as a function of voicing, because it is possible for speakers to control tongue height and jaw position independently. Clearly, more research will be necessary for one to understand better the relationship between the articulation of final stop voicing and *F1* offset frequency.

Final Remarks

We have seen that native Arabic speakers use vocalic duration to a small extent as a final consonant voicing cue in the English CVCs *pod* and *pot*. This finding conflicts with predictions of Crowther and Mann's phonemic vowel length hypothesis. One explanation for the prediction's failure is that the Arabic subjects' experience with vocalic duration as a cue to vowel identity in Arabic CVCs may have hindered their ability to learn to use vocalic duration for final consonant voicing within the same syllabic context in English. They are therefore unlike native English speakers, who seem able to use vocalic duration for two purposes: as a vowel cue, and as a voicing cue. The results of Experiments 2 and 3 suggest that the native Arabic subjects' apparent sensitivity to vocalic duration in Flege's (1984) *peas-peace* experiment may have resulted from task demands. Flege (1984) suggested that Arabic subjects may categorize stimuli in a *peas-peace* perception experiment on the basis of the Arabic vowel length distinction, rather than on the basis of voicing perception (which presumably is the basis used by native English speakers). Another explanation, rooted in terms of the perceptual anchor theory of context coding (Braida et al., 1984; Macmillan, 1987; Macmillan et al., 1987) and phoneme perception theory (Schouten & Van Hoesen, 1992; Van Hoesen & Schouten, 1992), provides an equally adequate account of performance differences between (1) Flege (1984) and Experiment 3 and (2) Experiment 2. These findings suggest that speech researchers need to consider carefully experimental design factors such as whether or not to use a factorial design; the number of stimuli to be employed in each condition; and the physical range and physical spacing between adjacent members along each dimension of the stimulus continuum (i.e., step size in frequency or duration). The perceptual anchor theory as well as the phoneme perception theory may be quite useful for testing the stability and structure of phoneme representations (or prototypes) for native and non-native English speakers. Poorly developed (or non-existent) prototypes may be implicated when a subject in a stop-consonant perception experiment seems to be highly influenced by the structure of the stimulus set (number of stimuli, range, step size, etc.). The application of these models suggests that any serious testing of sensitivity to speech cues should involve more extensive experimentation than is typical in speech perception experiments. Sub-

jects would have to be tested in fixed and roving discrimination, and in identification, and stimulus set factors would have to be manipulated, as well.

Native Arabic speakers use *F1* offset frequency as a voicing cue to distinguish the syllables *pod* and *pot*. This finding is agreement with the work of Crowther and Mann (1992), who showed that native Japanese and Mandarin speakers may be more similar to native English speakers in their use of *F1* offset than in their use of vocalic duration. The question of whether *F1* offset qualifies as a language-independent cue, as opposed to a cue that happens to be easy to learn, is open for further study. The present study offers another indication that the nature of the use of vocalic duration and that of the use of *F1* offset frequency may be fundamentally different.

REFERENCES

- BRADY, S. A., & DARWIN, C. J. (1978). Range effect in the perception of voicing. *Journal of the Acoustical Society of America*, **63**, 1556-1558.
- BRAIDA, L. D., LIM, J. S., BERLINER, J. E., DURLACH, N. I., RABINOWITZ, W. M., & PURKS, S. R. (1984). Intensity perception: XIII. Perceptual anchor model of context coding. *Journal of the Acoustical Society of America*, **76**, 722-731.
- CHEN, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, **22**, 129-159.
- CROWTHER, C. S. (1994). *The dependence between voicing and vowel identification when cued by the same spectral and temporal parameters*. (Manuscript in preparation.)
- CROWTHER, C. S., & MANN, V. (1992). Native language factors affecting use of vocalic cues to final consonant voicing in English. *Journal of the Acoustical Society of America*, **92**, 711-722.
- DIEHL, R. L., & KLUENDER, K. R. (1987). On the categorization of speech sounds. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 226-253). Cambridge: Cambridge University Press.
- FANT, G. (1962). Sound spectrography. In *Proceedings of the Fourth International Congress of Phonetic Sciences* (pp. 14-33). The Hague: Mouton De Gruyter.
- FLEGE, J. E. (1984). The effect of linguistic experience on Arabs' perception of the English /s/ vs. /z/ contrast. *Folia Linguistica*, **18**, 117-138.
- FLEGE, J. E. (1988). The development of skill in producing word-final English stops: Kinematic parameters. *Journal of the Acoustical Society of America*, **84**, 1639-1652.
- FLEGE, J. E. (1989). Chinese subjects' perception of the word-final English /t/-/d/ contrast: Performance before and after training. *Journal of the Acoustical Society of America*, **86**, 1684-1697.
- FLEGE, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*, **93**, 1589-1608.
- FLEGE, J. E., & HILLENBRAND, J. (1986). Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English. *Journal of the Acoustical Society of America*, **79**, 508-517.
- FLEGE, J. E., & PORT, R. (1981). Cross-language phonetic interference: Arabic to English. *Language & Speech*, **24**, 125-146.
- FLEGE, J. E., & WANG, C. (1989). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/-/d/ contrast. *Journal of Phonetics*, **17**, 299-315.
- FUJISAKI, H., NAKAMURA, K., & IMOTO, T. (1975). Auditory perception of duration of speech and non-speech stimuli. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 197-219). New York: Academic Press.
- GASS, S. (1984). Development of speech perception and speech production abilities in adult second language learners. *Applied Psycholinguistics*, **5**, 51-74.
- HILLENBRAND, J., INGRISANO, D. R., SMITH, B. L., & FLEGE, J. E. (1984). Perception of the voiced-voiceless contrast in syllable-final stops. *Journal of the Acoustical Society of America*, **76**, 18-26.
- KEATING, P. A. (1985). Universal phonetics and the organization of grammars. In V. A. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 115-132). Orlando, FL: Academic Press.
- KENT, R. D., & MOLL, K. L. (1969). Vocal-tract characteristics of the stop cognates. *Journal of the Acoustical Society of America*, **46**, 1549-1555.
- KLATT, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, **67**, 971-995.
- KLUENDER, K. R., DIEHL, R. L., & WRIGHT, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*, **16**, 153-169.
- LADEFOGED, P. (1982). *A course in phonetics* (2nd ed.). New York: Harcourt Brace Jovanovich.
- MACMILLAN, N. A. (1987). Beyond the categorical/continuous distinction: A psychophysical approach to processing modes. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 53-85). New York: Cambridge University Press.
- MACMILLAN, N. A., BRAIDA, L. D., & GOLDBERG, R. F. (1987). Central and peripheral processes in the perception of speech and nonspeech sounds. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 28-45). Boston: Martinus Nijhoff.
- MACMILLAN, N. A., GOLDBERG, R. F., & BRAIDA, L. D. (1988). Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua. *Journal of the Acoustical Society of America*, **84**, 1262-1280.
- MASSARO, D. W., & COHEN, M. M. (1976). The contribution of fundamental frequency and voice onset time to the /zi/ - /si/ distinction. *Journal of the Acoustical Society of America*, **60**, 704-717.
- MASSARO, D. W., & ODEN, G. C. (1980). Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America*, **67**, 996-1013.
- MERMELSTEIN, P. (1978a). Difference limens for formant frequencies of steady-state and consonant-bound vowels. *Journal of the Acoustical Society of America*, **63**, 572-580.
- MERMELSTEIN, P. (1978b). On the relationship between vowel and consonant identification when cued by the same acoustic information. *Perception & Psychophysics*, **23**, 331-336.
- MUNRO, M. J. (1990). Attention to spectral and temporal cues in vowel perception among native speakers of Arabic and English. *Journal of the Acoustical Society of America*, **88** (Suppl. 1), S53.
- NEAREY, T. M., & HOGAN, J. T. (1986). Phonological contrast in experimental phonetics: Relating distributions of production data to perceptual categorization curves. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology* (pp. 141-161). Orlando, FL: Academic Press.
- ODEN, G. C., & MASSARO, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, **85**, 172-191.
- PARKER, E. M., DIEHL, R. L., & KLUENDER, K. R. (1986). Trading relations in speech and nonspeech. *Perception & Psychophysics*, **39**, 129-142.
- RAPHAEL, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America*, **51**, 1296-1303.
- ROSEN, S. M. (1979). Range and frequency effects in consonant categorization. *Journal of Phonetics*, **7**, 393-402.
- SCHOUTEN, M. E. H., & VAN HESSEN, A. J. (1992). Modeling phoneme perception: I. Categorical perception. *Journal of the Acoustical Society of America*, **92**, 1841-1855.
- SHANNON, V., PALUMBO, R. L., & GRANDGENETT, M. (1988). IDENTIFY: A program for speech recognition, identification, and categorization experiments on PC compatible computers. *Journal of the Acoustical Society of America*, **83** (Suppl. 1), S16.
- SHELDON, A., & STRANGE, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, **3**, 243-261.
- STEVENS, K. N., KEYSER, S. J., & KAWASAKI, H. (1986). Toward a

- phonetic and phonological theory of redundant features. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 426-449). Hillsdale, NJ: Erlbaum.
- SUMMERS, W. V. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *Journal of the Acoustical Society of America*, **82**, 847-863.
- SUMMERS, W. V. (1988). F1 structure provides information for final-consonant voicing. *Journal of the Acoustical Society of America*, **84**, 485-492.
- VAN HESEN, A. J., & SCHOUTEN, M. E. H. (1992). Modeling phoneme perception: II. A model of stop consonant discrimination. *Journal of the Acoustical Society of America*, **92**, 1856-1868.
- WALSH, T., & PARKER, F. (1983). Vowel length and vowel transition cues to [+/- voice] in post-vocalic stops. *Journal of Phonetics*, **11**, 407-412.
- WESTBURY, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America*, **73**, 1322-1336.
- WHALEN, D. H. (1989). Vowel and consonant judgments are not independent when cued by the same information. *Perception & Psychophysics*, **46**, 284-292.
- WOLF, C. G. (1978). Voicing cues in English final stops. *Journal of Phonetics*, **6**, 299-309.

NOTE

1. It should be noted that this hypothesis is intended to describe the "natural" learning process, in which the speaker is not taught explicitly to lengthen vocalic segments before voiced stop consonants. If a speaker were taught the English vocalic duration voicing association, his/her productive mastery obviously would not necessarily imply perceptual mastery.

(Manuscript received October 23, 1992;
revision accepted for publication September 30, 1993.)