THE EFFECT OF LINGUISTIC EXPERIENCE ON ARABS' PERCEPTION OF THE ENGLISH /s/ vs. /z/ CONTRAST

JAMES EMIL FLEGE

This study addresses the issue of whether the perception of L2 speech sounds evolves during the course of naturalistic L2 acquisition by adults. Synthetic speech stimuli were presented to three subject groups for forced choice identification as "piece" or "peas": (1) monolingual speakers of American English; (2) relatively inexperienced Arab speakers of English; and (3) experienced Arabs. The experienced Arabs responded much like the Americans. They showed sharp increases in /z/ responses as vowel duration increased; as fricative duration decreased; and as vowel and fricative duration was inversely modified in the same CVC stimulus. The inexperienced Arabs resembled the American subjects on only 2 stimulus continua. They inconsistently labelled stimuli in which just final fricative duration was varied, probably because there is no temporal difference between word-final /s/ and /z/ in Arabic. It thus appears that the temporal difference distinguishing /s/ from /z/ in English but not Arabic words acquired perceptual salience for the more experienced of the two groups of Arab L2 learners.

Very few individuals are truly monolingual. Most people enter into regular contact with speakers of language varieties whose phonetic and phonological structure differs from their own native dialect. Second language (L2) learning represents just one important instance of such cross language (or dialect) contact between divergent sound systems (WEINREICH, 1953). The issue explored in this article is how such exposure to other language varieties might affect speech perception.

A phenomenon familiar to language teachers is the seeming inability of some L2 learners to distinguish L2 sounds which are not distinctively different in their own native language (L1). For example, the Japanese learner often confuses English /r/ and /l/ in both production and perception. The native Spanish speaker tends to confuse English /i/ vs. /j/. This is because, in both cases, the two L2 sounds (or near phonetic equivalents) are realizations of a single L1 phoneme. Researchers generally agree that such pronunciation patterns stem from acquired differences between L1 and L2 in language structure. However, recent speech perception research in-
icates that similar discriminative failures may occur before L1 acquisition commences.

WERKER, GILBERT, HUMPHRY and TEES (1981) report that adult Hindi speakers, but not English speaking adults, were able to discriminate naturally produced dental vs. retroflex and breathy voiced vs. voiceless aspirate dental Hindi stops. On the other hand, six-month-old infants being raised in an English speaking environment were able to discriminate both of these non-English stop contrasts, which is consistent with previous perception studies showing that pre-linguistic infants are able to discriminate many, if not all, of the speech sound contrasts occurring in natural human languages. In a longitudinal study, WERKER and TEES (1982) tested the ability of other infants to discriminate the contrast between dental vs. retroflex Hindi stops and velar vs. uvular glottalized stops found in Thompson (a Salish language). The infants discriminated both non-English sound contrasts at the age of 6 months but were no longer able to do so at 12 months of age, leading the authors to conclude that speech perceptual abilities may diminish at an early age in the absence of specific linguistic experience (cf. TREHUB, 1976).

Findings such as these raise the issue of the extent to which speech perceptual abilities develop according to an inborn maturational schedule without reference to environmental stimuli, and to what extent it depends on experience with the phonetic contrasts of specific languages for its development and/or maintenance (ASLIN and PISONI, 1980). In the context of L2 acquisition, it raises two questions. First, are the apparent decreases in discriminative ability noted for infants and adults due to neurological or experiential factors? Second, can diminution in speech perceptual abilities, when noted, be reversed during the course of naturalistic L2 acquisition?

One might infer from investigations of accentedness in adult L2 production (KRASHEK, SCARCELLA and LONG, 1982) that linguistic experience has comparatively little effect on phonetic representations after a “sensitive” or “critical” period for speech acquisition has passed (but cf. FLEGE, 1981). Certain aspects of the development of visual perception lend prima facie support to the notion that listeners’ attunement to L1 phonetic characteristics may in fact be irreversible (see ASLIN and PISONI, 1980, for an excellent discussion). If so, it is probably futile to attempt improvement of L2
comprehension, and possibly production, in the classroom or clinic.

However, such a conclusion appears unwarranted at the present time. First, although infants who have presumably not passed the "critical" period may evidence failure to discriminate phonetic contrasts not found in the ambient speech community, older children are believed by many to be quite capable of pronouncing foreign languages with an authentic accent. This suggests that reversible attentional and/or experiential factors are the source of at least the infants' difficulty. Second, gradual improvement in L2 production ability by adult L2 learners is frequently noted in instrumental studies of L2 production (Flege, 1980; Flege and Port, 1981). Such modifications in production may well be accompanied (or preceded by) changes in speech perceptual processes or phonetic representations. Third, several studies reviewed below provide direct evidence that speech perception evolves as a result of exposure to or training on new phonetic contrasts.

I. PREVIOUS RESEARCH IN SECOND LANGUAGE PERCEPTION

Previous research provides several indications of a change in speech perception resulting from exposure to new or different phonetic contrasts. Several studies have examined perception of the syllable-initial stop voicing contrast in English as a function of VOT. VOT, which refers to the relative timing of stop release and vowel onset, functions as a perceptual cue to the contrast between stop categories such as /b, d, g/ and /p, t, k/ in most (if not all) languages where stops are distinguished in production by VOT.

Caramazza, Yeni-Komshian, Zurif, and Carbonell (1973) examined stop consonant perception by monolingual and bilingual speakers of French and English in Canada. For monolingual French speakers the boundary between predominantly voiced (/b, d, g/) and voiceless (/p, t, k/) stop responses occurred at about +15 ms, as compared to about +30 ms for English monolinguals. For French—English bilinguals, on the other hand, the phoneme boundary occurred at an intermediate point (averaging +23 ms) along the VOT continuum for both French and English words. This suggests that the acquisition of L2 (English) affected their perception of L1 sounds as much as it did L2 sounds.
Williams (1980) also varied VOT in her study of adult bilinguals who learned their L2 (either Spanish or English) in early childhood and whose pronunciation of both L1 and L2 was demonstrably excellent. Most subjects responded in an English-like fashion, showing a sharp crossover from predominantly /ba/ to /pa/ responses at a point along the VOT continuum which approximated (but did not reach) the boundary of monolingual English subjects (+24 ms). They also showed a sharp increase in discriminability for stimuli straddling (i.e., on opposite sides of) the phoneme boundary. However, several subjects revealed what might be considered to be a “bilingual” pattern of perception: their identification and discrimination functions indicated sensitivity to the phonetic contrast distinguishing voiced from voiceless stops in Spanish, as well as to the phonetically different contrast distinguishing voiced from voiceless stops in English.

Williams' (1980) study also examined Puerto Rican children aged 8–10 years and 14–16 years. Both groups showed a gradual shift from Spanish-like to English-like patterns of perception. The crossover from predominantly /b/ to /p/ responses occurred at increasingly large VOT values as a function of increased experience with English (i.e., in the direction of the English boundary at +24 ms and away from the Spanish monolingual boundary at -4 ms). However, the children’s phoneme boundaries nonetheless remained intermediate in value to the boundaries of monolingual speakers of Spanish and English. Perhaps the major difference between groups was that only the older children — like some of the adult Spanish–English bilinguals mentioned earlier — showed evidence of heightened discriminability at both the Spanish and English stop category boundaries. Williams concluded that, in spite of the perceptual changes noted, even child L2 learners might never fully resemble monolingual speakers in terms of speech perception.

Other studies have examined whether Ss can be trained to perceive a new category of stops through laboratory training. Lisker (1970) presented endpoint stimuli of a synthetic VOT continuum (VOT values of +10 ms and +60 ms) to native speakers of Russian. Stimuli such as these will be heard by English speakers as /ba/ and /pa/, but both are heard as /pa/ by monolingual Russian speakers because the contrast between “short-lag” (e.g., +10 ms) and “long-lag” (e.g., +60 ms) stops is not linguistically distinctive in Russian. Discrimination training enabled a group of Russians to
assign different category labels to the two stimuli, although the Russian subjects did not divide the continuum as consistently as native English speakers.

Strange (reviewed in Strange and Jenkins, 1975) attempted to train English speakers on the prevoiced stop category found in languages such as Thai. Subjects heard CV syllables synthesized with VOT values ranging from $-150$ to $+150$ ms. Whereas Thai native speakers perceive three stop categories in such a continuum (voiced, voiceless unaspirated, and voiceless aspirated) native English speakers typically hear only two (a voiceless category, and a voiced category encompassing both “lead” and “short-lag” stops). Training of 5–7 hours, with either immediate or delayed feedback, resulted in only limited improvement in the discrimination of the “lead” from “short-lag” stimuli. At the same time, training did not generalize to other places of articulation, leading Strange to conclude that “changing the perception of VOT dimensions ... is not easily accomplished by techniques that involve several hours of practice” (Strange and Jenkins, 1975: 152).

However, results of a similar study (Pisoni, Aslin, Perey and Hennessy, 1982) provides evidence that native English speakers can learn a category cued by “lead” VOT values. Subjects trained simply by the alternating presentation of “good” examples of “mba”, “ba”, and “pa” ($-70$, $0$, and $+70$ ms VOT) rapidly learned to distinguish the lead (“mba”) from short-lag (“ba”) categories. The study was later replicated by McClasky, Pisoni and Carrell (1980), whose results also demonstrated generalization of training to a previously untrained VOT continuum. This finding probably will come as no surprise to linguists and foreign language teachers. However, its importance lies in empirically demonstrating that speech perceptual processes need not be irreversibly constrained by previous linguistic experience, and that relatively little training — comparable to what might reasonably be spent in a classroom — may effect such changes.

Several other studies relate to the perception of /r/ vs. /l/ by Japanese L2 learners. As is well known, native speakers of Japanese often confuse these sounds in production and perception because there is no /r/ vs. /l/ contrast in their native language (Price, 1981). Adult Japanese subjects differ dramatically from native English speaking subjects in discriminating or labelling synthetic speech stimuli which, for Americans, span a continuum from /r/ to /l/
Identification of naturally produced minimal pairs may be somewhat less difficult for Japanese native speakers (Mochizuki, 1981), but even then performance is still less consistent than for native English speakers.

A recent study by Mackain, Best, and Strange (1981) indicates that the ability of Japanese L2 learners to perceive English /r/ and /l/ may evolve during the course of naturalistic acquisition. Japanese learners with relatively great experience in English conversation perceived stimuli in an /r/-to-/l/ continuum like native English speakers, showing a "categorical" pattern of perception when labeling and discriminating stimuli in a "rock" to "lock" continuum (i.e., a sharp cross-over from "lock" to "rock" judgments, with increased discriminability for stimuli straddling the phoneme boundary). Japanese L2 learners with less experience speaking English, on the other hand, showed inconsistent responses both when discriminating and labeling the synthetic stimuli.

This study is important because it demonstrates that perception may evolve unaided. Strange and Dittman (1981) attempted to reproduce this same evolution in the laboratory, providing /r/ vs. /l/ discrimination training with immediate feedback to eight Japanese women. Discrimination improved slowly but steadily during 15 training sessions, and post-training identification improved dramatically as a result. However, transfer of training to another synthetic speech continuum and to naturally produced /r/ vs. /l/ minimal pairs was much less certain.

Finally, a study by Esendoorn (1980, 1981) indicates that Dutch university students, but not Dutch high school students, showed the same kind of sensitivity to vowel duration as native English speakers in judging word final stops as voiced or voiceless. Since earlier studies suggest little effect of vowel duration on stop voicing judgements in Dutch words, this finding suggests that with sufficient experience in English, temporal differences may become a perceptual cue to the word-final stop voicing contrast for L2 learners.

II. PERCEPTUAL LEARNING IN L2 ACQUISITION

Based on previous studies, four hypotheses regarding how L2 acquisition might affect speech perception can be formalized for cases in which cognate L1 and L2 sounds differ phonetically. First,
phonetic interference might be so thorough and persistent that no modification of phonetic representations occurs during L2 acquisition. L2 learners might simply go on producing and perceiving L2 sounds that differ phonetically from L1 cognates as if they were identical to their L1 cognates. This might explain the misperceptions of L2 words that sometimes occur, especially in non-optimal (e.g. noisy) listening conditions. However, while such a pattern might exist in the earliest stages of L2 acquisition, the evidence now existing suggests it does not persist (ELMAN, DIEHL and BUCHWALD, 1977; WILLIAMS, 1980; MACKAIN et al., 1980; cf. FLEGE and PORT, 1981).

Second, phonetic representations might be restructured to accommodate the divergent phonetic norms of L1 and L2 so that L2 learners come to perceive L1 and L2 sounds differently from monolingual speakers of either L1 or L2. Evidence of such “compromise” perceptual patterns (WILLIAMS, 1980) could take two forms. First, the boundary between categories could occur at values which are intermediate to those seen for monolingual speakers of L1 and L2 (e.g., some of the Spanish—English bilinguals in WILLIAMS’, 1980 study). Second, the crossover region (where stimuli are inconsistently labelled) might be widened to embrace the category boundaries found in L1 and L2. This pattern, represented by a more gradual slope in the identification function (or non-monotonic functions) is seen in child L1 acquisition studies (ZLATIN and KORNIGSKNECHT, 1976) and characterized the perception of stops by French—English bilinguals in the CARAMAZZA et al. (1973) study.

Third, the L2 learner might abandon the L1 pattern of perception in favor of the L2 pattern, as suggested by the behaviour of some of WILLIAMS’ (1980) Spanish—English bilingual subjects. This might occur in cases where the phonetic contrast in L2 is easier to discriminate for auditory reasons than the L1 phonetic contrast.

Fourth, experience with L2 might result in an “enriched” phonetic space in which phonetically different L1 and L2 cognate sounds have separate (or at least functionally separable) phonetic representations. This is what we might expect of a “true” bilingual: the ability to authentically produce L1 and L2 sounds, and to perceive sound contrasts in L1 and L2 like monolingual speaking subjects. However, only one of three existing experimental investigations have demonstrated distinct patterns of perception by bi-
linguals for L1 and L2 sounds (ELMAN, DIEHL and BUCHWALD, 1977). At the same time, no study that I know of has demonstrated a perfect conformity to L1 and L2 phonetic norms in the speech of bilinguals.

Inasmuch as it is influenced by linguistic experience, the exact nature of L2 speech perception is apt to be affected by factors such as age of acquisition, the nature of early learning and usage patterns, bilingual balance, and little understood psycho-social factors. It is likely that none of the hypotheses offered above will uniquely account for perceptual modification during, and as a result of, L2 acquisition. However, before any of the above hypotheses can be eliminated or more narrowly constrained, we must first know more about the perception of L2 sounds by language learners.

The goal of the present study was to provide further evidence concerning possible changes in speech perception during L2 acquisition. Previous studies have examined the VOT dimension in prevocalic stops, the /r–l/ contrast, and the effect of vowel duration on postvocalic stop voicing. This study focusses on the effect of temporal variables on the perception of postvocalic fricatives. To determine whether perception evolves, two groups of learners differentiated by experience in L2 are compared.

III. PHONETIC DIFFERENCES BETWEEN ARABIC AND ENGLISH

It has long been recognized that language background systematically affects how subjects respond to stimuli varied along a single phonetic dimension such as VOT (LISKER and ABRAMSON, 1970). In an attempt to emulate real speech, many recent perception experiments manipulate two or three acoustic dimensions orthogonally. A recurrent finding is that no single auditory-acoustic dimension equivalent to a “distinctive feature” cues a phonemic contrast (FLEGE, 1977). The net perceptual effect of a dimension which, in itself, might be sufficient to cue a phonemic contrast is often altered in the presence of variation in the stimulus value of another dimension. Such a “trading relation” is seen in the inverse pattern of temporal differences which in part cue the distinction between word final voiced and voiceless obstruents in English.
Production studies indicate that in English the closure (stricture) duration of voiceless obstruents exceeds that of voiced homorganic cognates, while vowels are substantially longer before voiced than voiceless obstruents. Correspondingly, studies show that English speaking subjects are sensitive to this inverse temporal variation in vowel and obstruent duration. Listeners tend to hear voiceless obstruents when obstruent duration is relatively long and the preceding vowel is relatively short. Within limits, the duration of one interval affects the potency of temporally manipulating the other.

Vowels in CVC citation forms may be as much as 70% longer before /z/ than /s/, although the absolute duration of vowels may be smaller, and the magnitude of the voicing effect on vowel duration either smaller or absent, in utterance-medial positions. Peterson and Lehiste (1960) report mean durations of 390 ms for vowels preceding /z/, compared to 269 ms for vowels before /s/. Raphael (1971) reports durations of 365 ms (pre-/z/) versus 218 ms (pre-/s/). Denes (1955) reports relatively shorter vowel durations ranging from 40 – 200 ms in the environment of /s/ and /z/.

In final position /s/ duration exceeds /z/ duration, at least in citation forms. For example, Raphael (1971) reports average values of 275 ms for /s/ compared to 195 ms for /z/. Denes (1955) reports values ranging from 200 – 350 ms for /s/ compared to only 100 – 180 ms for /z/.

In a perception study, Raphael (1971) varied vowels from 200 – 400 ms in a “deuce” to “dues” continuum. The phoneme boundary occurred at a vowel duration of 271 ms. Gruenenfelder (1979) varied vowel duration from 50 to 350 ms in a “cease” to “seize” continuum, reporting a boundary crossover at about 180 ms. Denes (1955) varied both vowel duration (from 50 to 200 ms) and fricative duration (from 50 to 250 ms) using hybrid synthetic-natural stimuli. Subjects heard voiceless fricatives (/yus/) when the ratio of final fricative to preceding vowel duration was greater than about 1.0; and a voiced fricative (/yuz/) for smaller C/V ratios. Derr and Massaro (1980) orthogonally varied vowel duration (from 40 – 200 ms) and fricative duration (from 70 – 210 ms) using synthetic five-formant stimuli. Both temporal manipulations produced significant shifts in voicing judgments (as /yus/ or /yuz/), although the endpoint stimuli were not consistently labelled. Soli (1982) edited natural tokens of “use” (noun) and “use” (verb) to produce a continuum of stimuli whose vowels ranged from 250 – 450 ms and whose fricatives ranged from 125 – 275 ms. He noted seemingly larger effects for variations in vowel duration and quality than for fricative duration. However, even when other perceptual cues were ambiguous, /z/ judgments increased from about 20% to 75% as final fricative duration was increased from 10 – 200 ms in a separate experiment.

Several studies show that Arabic does not possess the same inverse temporal differences distinguishing voiced from voiceless obstruents in the word final position of English words. Only relatively small and often non-significant effects of obstruent voicing on preceding vowel duration have been observed in Arabic-accented English. Similarly, a much smaller and sometimes non-significant temporal contrast between voiced and voiceless obstruents is noted in post-vocalic position (Fleiss, 1980; Fleiss and Port, 1981; Mittler, 1981; Port and Mittler, 1980).

Both non-English characteristics appear to be the direct result of phonetic interference from Arabic to English, for Arabic appears to lack the temporal correlates that distinguish voiced from voiceless
obstruents in English. MITLEB (1982) reports that Jordanians made vowels about 20 ms longer before /z/ than /s/ in English words. This contrast is somewhat larger than the 5 ms effect seen in their production of comparable Arabic words, but substantially smaller than the ca. 50 ms effect produced by native English speakers. At the same time, the native Arab speakers also produced a smaller contrast between /s/ and /z/ in Arabic (ca. 5 ms) and Arabic-accented English (ca. 15 ms) than did Americans producing English /s/ and /z/ (45 ms).

The nature of the relationship between developmental changes in speech perception and speech production are far from clear at the present time. However, since native speakers of Arabic do not seem to use the same temporal correlates to differentiate word-final /s/ vs. /z/ in either English or Arabic, it seems reasonable to ask whether the temporal manipulation of vowel and fricative duration will affect their perception of word final fricatives as /s/ or /z/. The following experiment addressed this question.

IV. METHODS

A. Stimulus Generation. A set of synthetic speech stimuli, illustrated in Fig. 1, was generated using a cascade-parallel software synthesizer (KLATT, 1980) that provides separate noise sources for vowels and fricatives and permits an updating of acoustic parameters every 5 ms.

Fig. 1. A schematized representation of the synthetic CVC stimuli (see text for explanation).
A 5-ms release burst at stimulus onset was generated by passing a turbulent noise source through the bypass channel of the parallel branch of the synthesizer. Aspiration noise between the release burst and onset of the "vowel" was simulated by passing a noise source (AH) through the cascade branch of the synthesizer and widening the bandwidths (BW) of the first three formants (F1 - F3). The starting frequencies of the aspirated formant transitions leading into the vowel were: F1 - 400 Hz; F2 - 1700 Hz; F3 - 2600 Hz. Fundamental frequency (F0) decreased linearly during the vowel from 128 to 110 Hz.

The relative amplitude of the vowel and fricative intervals, and the amplitude rise and fall times, were matched as closely as possible to the amplitude contours of natural productions of "piece" and "peas" displayed on spectrograms (Voiceprint, Model 700). Intensity of the periodic voicing source (AV) for the vowel reached its maximum intensity within 10 ms of onset. Center frequencies of the five formants were: F1 - 280, BW = 100; F2 - 2350, BW = 100; F3 - 3000, BW = 170; F4 - 3700, BW = 400; F5 - 4000, BW = 500. The amplitude of the vowel decreased abruptly during the final 15 ms, and somewhat less abruptly during the preceding 85 ms. The amplitude of the fricative noise source (AF) increased linearly for 85 ms, overlapping the vowel by 50 ms. Its amplitude decreased abruptly during the final 20 ms of the fricative interval, and somewhat less abruptly during the preceding 40 ms as a result of linearly interpolating the synthesizer parameter A6.

The range of values over which temporal parameters are varied has an important effect on how synthetic speech stimuli are identified (ROSEN, 1979), so it was important to select appropriate continuum endpoint values and step sizes. The fricative and vowel durations employed here were based on a careful consideration of the studies cited above and the speech of a native English speaker (JEF). Duration of the "vowel" and "fricative" portions of stimuli were modified by repeating or deleting 50-ms segments of these intervals as summarized in Table 1. Vowels ranged from 150-350 ms in five equal steps, fricatives from 100-300 ms, also in five equal steps.

Spectrograms of four of the synthetic stimuli illustrating extreme values of the temporal parameters are displayed in Fig. 2. Note that no low-frequency energy (voicing) is present during the word final fricative.

B. Pilot Study. A pilot was run to test the properties of the 25-member stimulus continuum. Stimuli were converted to analog form at a 10 kHz sampling rate using a 12-bit D/A converter, low-pass filtered at 4.9 kHz, and presented binaurally to subjects through matched and calibrated headphones (TDH-39) at a comfortable listening level of about 80 dB SPL(A). Presentation of each succeeding stimulus was paced to the slowest subject. Both stimulus presentation and response collection were run under the control of a small laboratory computer. Stimuli were randomized without replacement in each of 10 blocks. Six adult native English speakers with self-reported normal hearing identified stimuli in a forced choice task by pressing either a button marked "piece" or "peas."

The results displayed in Fig. 3 are in close agreement with previous studies in showing that both fricative and vowel duration together influence
Vowel and fricative duration parameters of the 25 stimuli used in the pilot study. The three digit numbers represent \( CV \) ratio (fricative over preceding vowel duration). The percentage of "peas" responses by subjects in the pilot experiment are also presented.

<table>
<thead>
<tr>
<th>Vowel Duration in ms</th>
<th>150</th>
<th>200</th>
<th>250</th>
<th>300</th>
<th>350</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fricative Duration, ms</td>
<td>100</td>
<td>150</td>
<td>200</td>
<td>250</td>
<td>300</td>
</tr>
<tr>
<td>1</td>
<td>2.00</td>
<td>1.50</td>
<td>1.00</td>
<td>0.67</td>
<td>0.50</td>
</tr>
<tr>
<td>0%</td>
<td>5%</td>
<td>3.3%</td>
<td>5%</td>
<td>27%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.33</td>
<td>1.00</td>
<td>0.75</td>
<td>0.50</td>
<td>0.40</td>
</tr>
<tr>
<td>8%</td>
<td>73%</td>
<td>92%</td>
<td>93%</td>
<td>93%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.00</td>
<td>0.75</td>
<td>0.60</td>
<td>0.40</td>
<td>0.30</td>
</tr>
<tr>
<td>55%</td>
<td>95%</td>
<td>98%</td>
<td>98%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.75</td>
<td>0.50</td>
<td>0.40</td>
<td>0.30</td>
<td>0.20</td>
</tr>
<tr>
<td>8%</td>
<td>73%</td>
<td>92%</td>
<td>93%</td>
<td>93%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.50</td>
<td>0.40</td>
<td>0.30</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td>5%</td>
<td>3.3%</td>
<td>5%</td>
<td>27%</td>
<td>44%</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.33</td>
<td>0.25</td>
<td>0.20</td>
<td>0.15</td>
<td>0.10</td>
</tr>
<tr>
<td>8%</td>
<td>73%</td>
<td>92%</td>
<td>93%</td>
<td>93%</td>
<td></td>
</tr>
</tbody>
</table>

Subjects judged an increasingly large proportion of stimuli as "peas" as vowel duration increased and fricative duration was decreased. Increasing vowel duration from 150—350 ms resulted in a net increase of 78% in /z/ responses (from 19 to 97%) in stimuli having fricatives 150—200 ms long. Decreasing fricative duration from 300—100 ms resulted in a net increase of 82% in /z/ responses (from 3 to 85%) when the preceding vowel was 200—300 ms long. All six subjects showed the same response pattern.

Following Denes (1955) the ratio of the final fricatives to that of preceding vowels was calculated for each of the 25 stimuli. In the bottom panel of Fig. 3 the proportion of "peas" responses is replotted as a function of the \( CV \) ratio. The present results are in substantial agreement with those of Denes (1955) and Port (1982) in showing a close relationship between voicing judgments and \( CV \) ratio. However, these results do not in themselves shed further light concerning whether the durations of these adjacent phonetic intervals are perceived independently or integrally (Derr and Massaro, 1981; Port, 1982).

C. Experimental Stimuli. Based on the pilot results, three subcontinua representing the entire range of temporal values employed in the pilot study were formed. Continuum I consisted of stimuli (1, 7, 13, 19, and 25; see Table 1) in which both vowel and fricative duration was varied in five steps. The endpoint stimuli of this continuum were identified as /z/ 0.0% (#1) and 100.0% (#25) in the pilot experiment. Continuum II consisted of stimuli (11—15) in which fricative duration was held constant at a neutral value of 200 ms and vowel duration was varied. The endpoints were identified as /z/ 0.0% (#11) and 88% (#15) of the time in the pilot.
Fig. 2. Spectrograms illustrating the extreme values of vowel and fricative duration parameters used in the synthetic speech stimuli.
Fig. 3. (Upper Panel) The mean percentage of responses as "peas" by 6 monolingual English subjects in the pilot experiment. Each data point is based on 60 judgements; (Lower Panel) the same data replotted as a function of C/V ratio (final fricative over preceding vowel duration).

Continuum III consisted of stimuli (3, 8, 13, 18, and 23) in which vowel duration was held constant at a neutral value of 250 ms and fricative duration was varied. The endpoint stimuli of this continuum were identified as /z/ 8.3% (≠3) and 98.0% (≠23) of the time.
It should be noted here that the possibility of extra-experimental bias is considerably reduced since /pls/ and /piz/ are not possible words in Arabic owing to the absence of a /p/ phoneme from Arabic.

D. Subjects. The three continua were presented to three subject groups differing in language background, all right-handers with normal hearing according to self report. One group consisted of 12 monolingual English speakers, American college students with a mean age of 20 years who participated for course credit.

There were two groups of native Arabic speakers who were maximally differentiated according to their English language ability. The first group, designated the “inexperienced” Arabs, consisted of five Saudi Arabian students (mean age: 28 years) who had arrived in the United States just two months prior to the experiment and had difficulty communicating in English. These individuals were enrolled in an English as a second language program at the time of the study and anticipated doing graduate work in Education. The “experienced” Arab group consisted of 6 somewhat older individuals (mean age: 33 yrs) who had lived in the U. S. for an average of 5.8 years (range: 2.8 to 18 years). They were specialists in linguistics or literature who were all highly proficient speakers of English in the author’s estimation. Subjects in this group came from several Middle Eastern countries: Egypt (2), Iraq (2), Jordan, and Saudi Arabia.

E. Procedures. Ten randomized blocks of each continuum were recorded on audio tape (AG-500, 7.5 ips) in a single order (I, II, III). Stimuli were presented binaurally to subjects at a comfortable listening level of about 80 dB SPL(A). The CVC stimuli were appended to a 600 ms synthetic carrier phrase “Show me...” to provide a stable temporal anchor for the voicing judgements. The interval between stimuli was fixed as 4.0 sec.

To fully insure that the inexperienced Arab subjects understood the task, instructions were read by the author in English, and in Arabic by an Arabic-speaking linguist. The meaning of the English words “piece” and “peas” was explained. Illustrations of these words appeared above the appropriate column on the answer sheet subjects used to record their responses. Subjects were told to make their best guess in cases of uncertainty.
V. RESULTS AND DISCUSSION

The mean percentage of stimuli identified as “peas” are plotted in Fig. 4. Because of the small number of Arab subjects available, the identification functions will be discussed here only in qualitative terms without statistical analyses.

Subjects in all three groups showed a clear crossover from predominantly /z/ to /s/ responses as vowel duration decreased and fricative increased inversely in Continuum I. For the monolingual Americans, /z/ responses increased from 3% to 95%. For the experienced Arabs, /z/ responses increased from 0% to 100%; and for the experienced Arabs, /z/ responses increased from 16% to 86%. Individual response patterns conformed to the group function with the exception of one subject in both the “inexperienced” Arab group and the American group.

Both the American and the inexperienced Arabs showed a clear crossover from /s/ to /z/ judgments as just vowel duration was increased in Continuum II. All subjects without exception showed a response pattern equivalent to the group function. Data is unavailable for the experienced Arabs.

![Fig. 4. The mean percentage of responses as “peas” by three speaker groups for continua in which: (a) vowel and fricative duration were inversely varied in five steps; (b) just vowel duration was varied; and (c) just final fricative duration was varied.](image)
We see evidence of a difference between subject groups for Continuum III, in which just final fricative duration was varied. The grouped data displayed in Fig. 4 suggests that the inexperienced Arabs did not use fricative duration as a perceptual cue to the contrast between /z/ and /s/. The experienced Arabs, on the other hand, resembled the American subjects in identifying stimuli with relatively short fricative intervals as /z/. Only 1 of the inexperienced Arabs showed an increase in /s/ responses of 50% or more as fricative duration increased from 100 to 300 ms. However, 4 of the 6 experienced Arabs, and 11 of the 12 Americans showed such a crossover from /z/ to /s/ as the word-final fricative increased in duration. This finding is reminiscent of the Mackain et al. (1980) study in which experienced but not inexperienced Japanese L2 learners perceived the /r−l/ distinction in an English-like fashion. From this study it appears that the experienced Arabs, as a result of their experience with English, became sensitive to the temporal differences which in part distinguish English /s/ and /z/ in word-final position.

It is somewhat surprising that the inexperienced Arabs did not respond to the variation in final fricative duration. As noted earlier, there is no temporal difference between Arabic /s/ and /z/. However, there is a quantity distinction between /s/ vs /ss/, /z/ vs /zz/, and /i/ vs /ii/ in Arabic (Al-ani, 1970). So why were the inexperienced Arabs generally unable to use fricative duration as a perceptual cue to the voicing contrast when, in the previous condition, they had unambiguously categorized stimuli as “piece” or “peas” based on variations in vowel duration?

It is possible that even though /pis/ and /piz/ are not possible Arabic words, these subjects judged the stimuli in continuum II on the basis of whether they perceived the vowel as phonemically long or short. Indirect support for this can be found in Mitlleb’s (1981a) study of English L2 production. Arab learners were found to produce a much greater duration distinction between English tense and lax vowels than native English speakers. Americans made /i/ only 14% longer than /i/, whereas Arab speakers of English produced a 46% difference (123 vs 84 ms), comparable to their 53% difference between long /ii/ and short /i/ in Arabic words. Perhaps some Arab L2 learners interpret the distinction between tense vs lax vowel pairs in English (such as /l−l/) as a quantity distinction rather than primarily as a vowel quality difference like native English speakers.
However, this will not explain why the inexperienced Arabs did not use a similar strategy when labelling stimuli that differed only in fricative duration. AL-ANI (1970) reports temporal values of 100—180 ms for single fricatives as against 250—350 ms for geminate fricatives, so the range of values in the synthetic fricatives (100—300 ms) should have been sufficient to cue a quantity distinction if this strategy were being employed.

A possible clue is provided in a recent discussion by ELIASSON (1982). Swedish is analyzed as having two basic quantity patterns in stressed syllables, /VC/ and /VCC/. Long vowels in the surface form [V: C] are derived from an underlying short vowel by a vowel lengthening rule. Geminate consonants in surface forms such as [VC:] are derived by a consonant lengthening rule. The [VC:] pronunciation by some Swedes of Finnish words having the surface form [VC] suggests the possibility that only vowel quantity differences exist underlyingly in Swedish, with the surface quantity distinction in consonants derived by rule.

Although I have no such evidence concerning Arabic quantity patterns, such an analysis might explain the present results if it were found to hold for Arabic. If the temporal differences between vowels but not following consonants is linguistically relevant in Arabic words, the Arab L2 learner might be expected to attend primarily to vowel duration distinctions in L2 words, at least at the very beginning. Only with experience would the Arab learner establish that the duration of final obstruents varies non-redundantly in English.

A more easily tested hypothesis is that the inexperienced Arabs' failure to exploit fricative duration as a cue to voicing has an auditory basis. In an accent detection task JOHANSSON (1975) found greater sensitivity to variations in vowel duration than in final stop duration. ELIASSON (1982) suggests that, owing to intensity differences between consonants and vowels, vowel duration differences are "more important than differences in consonant duration" for the perception of quantity relations in Swedish syllables (1982, p. 192). Based on experimental evidence, SOLI (1982) observes that fricative duration is a relatively "weaker" cue to voicing than preceding vowel duration for native English speakers.

Perhaps Arabs, and language learners in general, tend to exploit "strong" perceptual cues that are auditorily easy to perceive before they begin attending to "weaker" cues that are less easily perceivable.
At present relatively little is known either about how adult language learners, as in the present study, or child L1 learners (ROBSON, MORRONGIELLO, BEST and CLIFTON, 1982) integrate the multiple acoustic cues that specify phonetic category information in natural speech. Although the experienced Arabs labelled stimuli in the three "piece-peas" continua like American subjects, we cannot be certain that they actually perceived the synthetic stimuli just like native English speaking subjects. Native English speakers are thought to base word final voicing judgments on the ratio of consonant to preceding vowel duration (PORT, 1982), or by comparing the weighted temporal value of a final obstruent and that of a preceding vowel to duration values specified separately for each interval as part of category prototypes (DERR and MASSARO, 1980). It is possible that, unlike English-speaking subjects, the Arab subjects perceived vowel and fricative duration in terms of Arabic quantity patterns rather than in terms of their correlation with the voicing characteristic of the word-final fricative.

Taken together within several previous studies of L2 perception (MACKAIN et al., 1981; ELSENDOORN, 1980, 1981; WILLIAMS, 1980) the present findings indicate that adult learners of English gradually come to perceive English sounds like native English speakers. This suggests that the effect of prior linguistic experience is not unmodifiable by subsequent experience. Further investigation is clearly needed to further determine the nature of the experience that is necessary and sufficient to induce such changes in L2 speech perception. Based on previous studies, one tentative conclusion is that non-meaningful training in laboratory conditions may not be as effective in modifying speech perception as hearing and speaking L2 in naturalistic conditions.

ACKNOWLEDGMENTS

The author would like to thank Drs R. Port, D. Kewley-Port, T. Gruenenfelder, and T. Carrell for advice concerning speech synthesis; Drs D. Pisoni and G. Weismer for the use of laboratory facilities; Dr. F. Mitlob for assistance in locating the Arabic speaking subjects; and Mr. J. Forshee for his valuable software support.

Address of the author: James Emil Flege

The University of Alabama in Birmingham
Medical Center, Dept. of Biocommunication
Birmingham AL 35289  U. S. A.
REFERENCES


DENES, P. 1955.

"The contribution of vowel duration, F0 contour and fricative duration cues to the /s/-/s/ distinction". Perception & Psychophysics 27. 51-59.

EELGISSON, S. 1982.
"Transfer as evidence for phonological solutions". Studia Anglica Posnaniensia 14. 185-196.


"Phonetic approximation in second language acquisition". Language Learning 30. 117-134.

FLEGE, J. 1981.
"The phonological basis of foreign accent: A hypothesis". TESOL Quarterly 15. 443-455.

"Cross-language phonetic interference: Arabic to English". Language and Speech. 24. 126-146.


JOHANSSON, S. 1975.

KLATT, D. 1980.

KRASHEEN, S., R. SCARCELLA, and M. LONG. 1982.


