Interlingual identification and the role of foreign language experience in L2 vowel perception

OCKE-SCHWEN BOHN and JAMES EMIL FLEGE

University of Alabama, Birmingham

ADDRESS FOR CORRESPONDENCE
O.-S. Bohn, Englisches Seminar der Universität, Olshausenstr. 40, D-2300 Kiel 1,
Federal Republic of Germany

ABSTRACT
This article examines the perception of four English vowels (/i, ɪ, ɛ, æ/) by adult native speakers of German. From the standpoint of German, it appears that English /i, ɪ, ɛ/ are perceptually similar, if not identical, to German /i, ɪ, ɛ/, whereas /æ/ is a "new" vowel for German learners of English. The role of foreign language experience in the perception of second language vowels was examined through labeling responses to members of synthetic continua (beat-bit, bet-bat) in which vowel duration and spectrum were varied factorily. The subjects were relatively experienced and inexperienced second language (L2) learners and a monolingual English control group. The results suggest that L2 experience did not affect perception for the continuum with the two "similar" vowels /i/ and /ɪ/. However, for the continuum involving the "new" vowel /æ/, the experienced Germans more closely resembled the native English speakers than the inexperienced Germans. The predominant use of duration cues in differentiating the English /ɛ/-/æ/ contrast by the inexperienced Germans suggested that when spectral cues are insufficient to differentiate an L2 vowel contrast, duration will be used.

The hypothesis tested in the present study was derived from a model of second language (L2) speech learning that distinguishes between identical, similar, and new sounds in a cross-language context (see Flege, 1987, 1988; Flege & Hillenbrand, 1984). Identical sounds (e.g., English and German /m/) are unlikely to pose a learning problem, but similar sounds like English and German /u/ might offer persistent, if perhaps only subtle, problems for L2 learners. The hypothesis being tested states that adult L2 learners will substitute a sound from their first language (L1) for a similar L2 sound even after extended L2 exposure because of equivalence classification. On the other hand, the model hypothesizes that for new L2 sounds (e.g., English /ɜ/ for L1 German learners) that exceed a yet-to-be-specified "crucial similarity measure" (Wode, 1977), learners will eventually establish new phonetic categories as a result of phonetic learning that is unimpeded by equivalence classification.

Although the sounds of German and English have been examined thor-
oughly, it is far from certain how German learners perceive English vowels. The first part of the present study explores the relationship between English /i, ɪ, e, æ/ and German vowels. Based on the conclusions of the first part, the second part tests the hypothesis in an identification experiment that presented synthetic beat-bit and bet-bat continua to two groups of native German listeners differing in experience with English.

RELATIONSHIP BETWEEN ENGLISH /i, ɪ, e, æ/ AND GERMAN VOWELS

Previous research

A review of the literature does not clearly indicate whether the /i/ and /ɪ/ vowels of English are perceptually similar or identical to German /i/ and /ɪ/. The /e/ vowels of German and English are probably identical, while English /æ/ can be regarded as a new sound for speakers of Standard German.

In both English and German, the high front vowels /i/ and /ɪ/ differ in terms of spectral quality and duration. For native English listeners, the /i/-/ɪ/ contrast is cued primarily by spectral differences, while duration is a less important cue (Stevens, 1959a, 1959b). The situation for German is less clear. Perceptual studies have shown that German listeners tend to differentiate between German vowel pairs from the upper portion of the vowel area primarily on the basis of spectral differences (Sendlmeier, 1981; Wängler & Weiss, 1975; Weiss, 1972, 1976, 1978), but the relative importance of temporal cues may depend on the individual listener and/or the native dialect. Individuals from northern Germany may rely more on spectral than duration differences, whereas the opposite tends to be true of listeners from southern Germany (Weiss, 1972, 1976). Some acoustic studies have supported the auditory impression (Jones, 1960; Moulton, 1962) that German /i/ and /ɪ/ are somewhat more peripheral than English /i/ and /ɪ/ (cf. Delattre, 1964), but Disner's (1983) acoustic comparison of English /i/ and German /i/ revealed no significant differences. Moreover, Delattre's (1964) observation that short vowels such as /ɪ/ are more central in English and more open in German than long vowels such as /i/ does not seem to be true of all German dialects (cf. Ivonen, 1987).

Like English /i/ and /ɪ/, English /e/ also has a readily identifiable counterpart in German. The /e/ of both languages is short and nominally lax. Perceptual studies (Sendlmeier, 1981; Stevens, 1959a, 1959b; Weiss, 1972, 1976) indicate that short duration is an important cue to the identification of /e/ by native listeners of both English and German. No information on acoustic differences between English and German /e/ is available. Since English /e/ is not mentioned as a problem in textbooks for German learners of English, English and German /e/ may in fact be identical.

The vowel /æ/ may be regarded as new by German learners of English. Standard German has no long vowel in the low front portion of the vowel space. In fact, the entire inventory of German monophthongs consists of pairs governed by “the principle of long-close vs. short-open” (Weiss, 1972, p. 634), so that a long vowel like /æ/ that is low and front runs counter to the
organizing principle of German vowels. For most speakers of German, long and front implies close.

The identification of English /i, i, e, æ/ in terms of German response categories

The aim of this preliminary experiment was to examine the conclusions reached in the literature review. It sought to determine how native German speakers identify the English vowels /i, i, e, æ/, using categories of their L1, and to obtain confidence judgments from each subject on their own categorizations.

METHOD

Stimuli

Three tokens each of the English words beat, bit, bet, and bat produced by 2 monolingual native speakers of American English were drawn as representative examples from data collected for a previous study (Bohn & Flege, 1989). The 2 talkers (of 10) were a 35-year-old female and a 21-year-old male from the southeastern United States. They produced English consonant–vowel–consonant (CVC) words in the carrier phrase I will say ___. The CVC words were digitized at 10 kHz, amplitude-normalized, and stored on disk for acoustic analyses. Five separate randomizations of the twelve tokens by each talker (four words × three repetitions) were generated with an inter-stimulus interval of 4.0 sec and an inter-block interval of 10.0 sec.

Table 1 presents the acoustic characteristics of the vowel portions of the stimuli. The formant frequencies were measured using linear predictive coding (LPC) analysis by placing a 25.6 ms hamming window at the acoustic midpoint of the vowel. Vowel durations were measured using a waveform editor. Each of the stimuli was identified with 100% accuracy as beat, bit, bet, and bat, respectively, in a related experiment (see Bohn & Flege, forthcoming) by a panel of adult monolingual native speakers of American English.

Subjects

The English words were presented to 11 native speakers of German (6 females, 5 males) who participated as unpaid volunteers. The subjects (mean age: 31.7 years) had normal hearing according to self-report. They were natives of northern Germany and had spent all or most of their lives in the Kiel area of Schleswig-Holstein. Practical considerations made it impossible to get strictly monolingual subjects. A detailed language background questionnaire revealed that all subjects had studied English for at least 6 years in school, and 3 had spent more than a few weeks in an English-speaking environment. In addition, some subjects reported familiarity with Low German.
Table 1. Acoustic characteristics of the vowel portions of the stimuli in the German category listening experiment, produced by a male (M) and a female (F) native American English speaker

<table>
<thead>
<tr>
<th>Talker</th>
<th>Word</th>
<th>Duration</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F0</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>beat</td>
<td>131</td>
<td>263</td>
<td>2523</td>
<td>3171</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(18)</td>
<td>(48)</td>
<td>(45)</td>
<td>(16)</td>
<td>(7)</td>
</tr>
<tr>
<td>M</td>
<td>bit</td>
<td>108</td>
<td>438</td>
<td>1929</td>
<td>2739</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(44)</td>
<td>(16)</td>
<td>(45)</td>
<td>(2)</td>
<td>(4)</td>
</tr>
<tr>
<td>M</td>
<td>bet</td>
<td>116</td>
<td>621</td>
<td>1821</td>
<td>2643</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(25)</td>
<td>(24)</td>
<td>(26)</td>
<td>(51)</td>
<td>(45)</td>
</tr>
<tr>
<td>M</td>
<td>bat</td>
<td>182</td>
<td>711</td>
<td>1926</td>
<td>2604</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(45)</td>
<td>(13)</td>
<td>(19)</td>
<td>(130)</td>
<td>(7)</td>
</tr>
<tr>
<td>F</td>
<td>beat</td>
<td>125</td>
<td>329</td>
<td>2698</td>
<td>3219</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(36)</td>
<td>(54)</td>
<td>(52)</td>
<td>(186)</td>
<td>(8)</td>
</tr>
<tr>
<td>F</td>
<td>bit</td>
<td>96</td>
<td>415</td>
<td>2096</td>
<td>2861</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(26)</td>
<td>(27)</td>
<td>(127)</td>
<td>(159)</td>
<td>(6)</td>
</tr>
<tr>
<td>F</td>
<td>bet</td>
<td>145</td>
<td>530</td>
<td>2043</td>
<td>2856</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(20)</td>
<td>(30)</td>
<td>(63)</td>
<td>(35)</td>
<td>(3)</td>
</tr>
<tr>
<td>F</td>
<td>bat</td>
<td>195</td>
<td>656</td>
<td>2062</td>
<td>2866</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4)</td>
<td>(58)</td>
<td>(64)</td>
<td>(146)</td>
<td>(5)</td>
</tr>
</tbody>
</table>

Note: Mean values for three tokens each, duration in ms, formant frequency in Hertz, parentheses indicate ranges.

Procedure

The approximately 30-min experiment was conducted in a quiet room at each subjects' home in the Kiel area. The stimuli were presented binaurally (Realistic LV 10) using a portable tape recorder (Marantz PMD 420). The two talkers were presented in counterbalanced order to the listeners, who were told that they would hear words spoken by two Americans. In the first part of the experiment they were to circle on an answer sheet the German word that corresponded most closely to the vowel in the English word or to circle N if the English vowel was not represented by one of the German words listed: *biet*, *bitt*, *bett*, *bäht*, and *bert*. The order of alternatives on the response sheet was counterbalanced across listeners.³

The purpose of the second part of the experiment was to elicit confidence judgments of the listeners' own categorical judgments. The subjects heard the stimuli a second time and were told to indicate, on a line next to the words they had chosen in the first part, how confident they were, using a 5-point scale from *quite sure* to *uncertain*.

As part of the instructions, the subjects were told to base their choice among the five word alternatives on the answer sheet on their own pronunciation of these words. In order to remind them of this, the subjects were asked to read one row of word alternatives from the answer sheet aloud three times before listening to the stimuli. The readings were recorded for later
Table 2. Acoustic characteristics of the vowel portions of the German words produced by 11 German speakers in the German category listening experiment

<table>
<thead>
<tr>
<th>Word</th>
<th>Duration</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Males</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>biet</td>
<td>142</td>
<td>260</td>
<td>2062</td>
<td>2943</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>(61)</td>
<td>(14)</td>
<td>(219)</td>
<td>(204)</td>
<td>(18)</td>
</tr>
<tr>
<td>bitt</td>
<td>72</td>
<td>395</td>
<td>1730</td>
<td>2486</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>(17)</td>
<td>(20)</td>
<td>(120)</td>
<td>(197)</td>
<td>(27)</td>
</tr>
<tr>
<td>bett</td>
<td>92</td>
<td>529</td>
<td>1604</td>
<td>2422</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td>(20)</td>
<td>(13)</td>
<td>(157)</td>
<td>(219)</td>
<td>(22)</td>
</tr>
<tr>
<td>bäht</td>
<td>233</td>
<td>518</td>
<td>1693</td>
<td>2387</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>(67)</td>
<td>(52)</td>
<td>(120)</td>
<td>(224)</td>
<td>(22)</td>
</tr>
<tr>
<td>bert</td>
<td>222</td>
<td>584</td>
<td>1486</td>
<td>2337</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>(32)</td>
<td>(32)</td>
<td>(126)</td>
<td>(126)</td>
<td>(15)</td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>biet</td>
<td>148</td>
<td>317</td>
<td>2582</td>
<td>3417</td>
<td>222</td>
</tr>
<tr>
<td></td>
<td>(26)</td>
<td>(23)</td>
<td>(75)</td>
<td>(255)</td>
<td>(35)</td>
</tr>
<tr>
<td>bitt</td>
<td>88</td>
<td>453</td>
<td>2113</td>
<td>2857</td>
<td>215</td>
</tr>
<tr>
<td></td>
<td>(23)</td>
<td>(33)</td>
<td>(70)</td>
<td>(111)</td>
<td>(34)</td>
</tr>
<tr>
<td>bett</td>
<td>123</td>
<td>606</td>
<td>2016</td>
<td>2858</td>
<td>195</td>
</tr>
<tr>
<td></td>
<td>(21)</td>
<td>(64)</td>
<td>(76)</td>
<td>(149)</td>
<td>(16)</td>
</tr>
<tr>
<td>bäht</td>
<td>233</td>
<td>645</td>
<td>2134</td>
<td>2945</td>
<td>182</td>
</tr>
<tr>
<td></td>
<td>(40)</td>
<td>(94)</td>
<td>(154)</td>
<td>(94)</td>
<td>(15)</td>
</tr>
<tr>
<td>bert</td>
<td>217</td>
<td>721</td>
<td>1787</td>
<td>2795</td>
<td>189</td>
</tr>
<tr>
<td></td>
<td>(36)</td>
<td>(119)</td>
<td>(202)</td>
<td>(140)</td>
<td>(24)</td>
</tr>
</tbody>
</table>

*Note:* Mean values based on three repetitions by 5 males and 6 females each, duration in ms, formant frequency in Hz, parentheses indicate ± one standard deviation.

Instrumental analyses using the same taperecorder as for the listening experiment (Marantz PMD 420) and a stereo electret condenser microphone (Sony Model ECM-939LT).

Table 2 presents the acoustic characteristics of the vowel portions of the German words produced by the 5 male and 6 female Germans who listened to the English words. The acoustic measurements were made using the same procedures as for the English stimuli.

The 240 responses (2 Runs × 4 bVt Words × 3 Tokens × 2 Talkers × 5 randomizations) from each subject were tabulated in response matrices. Mean confidence scores were calculated for the responses to each stimulus category (beat, bit, bet, and bat) in terms of each response category (biet, bitt, bett, bäht, bert, and N). These scores were multiplied by a constant so that the maximum possible confidence score was 100.
Table 3. Frequency (top panel, in percent) and confidence judgments (bottom panel, maximum = 100) for categorizations of English bVt words in terms of German response categories (broken line indicates different responses for the male [M] and the female [F] talker)

<table>
<thead>
<tr>
<th>German Response Categories</th>
<th>BIET /i/</th>
<th>BITT /i/</th>
<th>BETT /e/</th>
<th>BÄHT /ɛ:/</th>
<th>BERT /ɛ:/ =/æ/</th>
<th>N*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Frequency</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>beat /i/</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bit /i/</td>
<td>99.7</td>
<td>98.2 (M)</td>
<td>1.2 (M)</td>
<td>0.6 (M)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bet /e/</td>
<td>4.8 (F)</td>
<td>60.6 (F)</td>
<td>10.9 (F)</td>
<td>23.0 (F)</td>
<td>0.6 (F)</td>
<td></td>
</tr>
<tr>
<td>bat /æ/</td>
<td>0.9</td>
<td>40.6</td>
<td>56.7</td>
<td>1.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Confidence judgments</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>beat /i/</td>
<td>99.2</td>
<td>98.3</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bit /i/</td>
<td></td>
<td>87.8 (M)</td>
<td>0.4 (M)</td>
<td>0.3 (M)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bet /e/</td>
<td>2.8 (F)</td>
<td>46.8 (F)</td>
<td>4.8 (F)</td>
<td>10.8 (F)</td>
<td>0.1 (F)</td>
<td></td>
</tr>
<tr>
<td>bat /æ/</td>
<td>0.4</td>
<td>27.9</td>
<td>35.0</td>
<td>0.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*N = None of the word categories.

Results

Table 3 shows that responses to beat and bit on the one hand and to bet and bat on the other were quite different, both in terms of the scatter among German categories identified with the English words, and in terms of the German subjects' confidence in their own judgments.

Overall, the results support the assumption that native German listeners perceive English /i/ and /i/ as equivalent to German /i/ and /i/1. The subjects unambiguously identified beat using biel and bit using bitt, and they indicated a high degree of confidence for these two cross-language identifications. Their judgments were probably based on the spectral properties of the stimuli since the durations of the vowel portions in beat and bit overlapped (cf. Table 1).

Results for bet varied according to the talker, but bett was the preferred response for both. While the listeners were quite confident about their identification of one (the male) talker’s bet tokens with bett, the other (the female) talker's bet tokens were not as uniquely matched with bett, and the listeners were not very confident about their judgments. The bert and bāht categories were also used as responses. The preferred responses to the bat tokens were bert and bāht. Although bert and bāht contain the lowest long front vowels in the listeners' dialect, they were not confident in using these categories.

The interpretation of the perceptual results may be aided by an acoustic
Figure 1. Distribution of English vowels (produced by 2 American talkers) and German vowels (produced by 11 speakers from northern Germany) in the Bark difference space. Ellipses with orthographic labels (\langle IE \rangle, \langle I \rangle, \langle E \rangle, \langle AH \rangle, \langle ER \rangle) represent the German vowels in biet, bitt, bett, baht, and bert, respectively. Phonetic symbols with subscripts (M = male, F = female) are located at the mean of three productions each of the English words beat, bit, bet, and bat.

comparison of the English stimuli to the German subjects' productions of the response categories. For this purpose, the frequency values given in Tables 1 and 2 were converted into Barks to obtain Bark difference scores, which normalized the gender differences in the data. Figure 1 presents the production data from the English talkers and from the German listeners in the Bark difference space defined by the Bark1–Bark0 and Bark2–Bark1 dimensions which have been shown to separate front vowels optimally (Syrdal, 1985; Syrdal & Gopal, 1986). These dimensions will be referred to here as "vowel height" and "frontness-backness," respectively.

The ellipses in Figure 1 represent 95% confidence levels based on the two principal components of variation for the vowels produced by the German subjects. Because of the highly variable pronunciation of bæht and bert, orthographic rather than phonetic symbols are used to label the ellipses. Phonetic symbols with subscripts (M for the male and F for the female talker) indicate the location of the mean values for the vowels in beat, bit, bet, and bat for the 2 American English talkers.
Two reservations have to be made in using the acoustic comparison shown in Figure 1 to explain the perceptual results. First, the German words were produced in isolation, whereas the English stimuli were edited from the end of utterances. Second, the Bark difference scores plotted in Figure 1 and durations given in Table 1 and 2 are important, but are not the only determinants of perceived vowel identity.

Figure 1 shows that the German subjects’ [i], [ɪ], and [e] tokens (as in biet, bitt, and bett, respectively) cover relatively small areas. The corresponding English vowels as produced by the two American talkers fall within these regions, except for the female talker’s [i], which is slightly lower than the German subjects’ [i]. The ellipses representing the German subjects’ vowels in bāht and bert cover relatively large areas, indicating considerable variation in the pronunciation of these words across subjects. The [e:] pronunciations for (āh) obtained in this study are likely to be exceptional because bāht is derived from an onomatopoeic expression. Northern Germans usually pronounce (āh) as [ː], and the use of [e:] for (āh) marks the speaker as coming from a different dialect area or as being hypercorrect. The pronunciation of (er) is also highly variable, probably because it is a strong social and regional indicator. Low variants for (er), as observed for some German subjects in the present study, are stereotypical of a Kiel accent and may be subject to social stigmatization. In the type of representation chosen here, the English talkers’ [æ] tokens are located at or close to the periphery of the German listeners’ vowels in bāht and bert.

DISCUSSION

The results for English /ɪ/ and /ɜ/ (in beat and bit) were to be expected from what is known about English and German vowels. More specifically, they are in agreement with the acoustic comparison of the English stimuli to the German subjects’ productions of biet and bitt in Figure 1. The fact that one English talker’s [i] was lower than the German subjects’ [i] did not result in low confidence scores, probably because there is no vowel category that competes with /i/ in the high front area of the vowel space. While this result does not tell us whether English /i/ and /ɪ/ are identical or merely similar to their German counterparts, it suggests that native German speakers treat these vowels as equivalent to German /i/ and /ɪ/. The fact that the vowels in beat and bit overlapped in duration indicates that the spectral information contained in these stimuli was sufficient to cue the perception of /i/ and /ɪ/ by native listeners from northern Germany.

The results for English /e/ and /æ/ (in bet and bat), on the other hand, suggest that these vowels do not have clear counterparts in German. This is somewhat surprising for /e/ since no learning problems of German speakers have been reported for English /e/. The one expectation is Oakeshott-Taylor (1976), who found that German learners of English tended to perceive (British) English /e/ as /æ/. Oakeshott-Taylor related this perceptual behavior to his German subjects’ intended production of English /e/, which was fre-
quently heard as /i/ by native English listeners. Apparently, the native /e/ prototype of his L1 German subjects was closer than the native English listeners' /e/, which also led his L1 German listeners to perceive English /e/ as the opener and longer vowel /æ/. Two interpretations for this are possible. Oakeshott-Taylor suggested that experience with English may have led his subjects to increase the distance (in perception and in production) between English /e/ and /e/, pushing their English /e/ higher in the vowel space because of system pressure. An alternative explanation assumes that German /e/ is closer and shorter than English /e/, and that Oakeshott-Taylor's subjects may have used their German /e/ in the production and perception experiments.

The perceptual findings of the present experiment, in which one American talker's bet was frequently identified with German bäht or bert (which have lower and longer vowels than bet) suggest that German /e/ is closer and shorter than English /e/. However, the acoustic comparison in Figure 1 shows that the English stimuli fall within an area covered by the German subjects' /e/ category. This suggests that the perceptual results for English bet were primarily due to the duration of the vowel portions of the stimuli. The durations in the three bet tokens that were frequently identified with bäht or bert were 133, 150, and 153 ms, whereas the durations of the vowel in bet produced by the German subjects were mostly shorter (cf. Table 2). The perceptual result of the present experiment is consistent with Weiss' (1976) observation that short duration (< 120 ms) is essential to the identification of /e/ by native Germans.

The lowest confidence ratings were obtained for the bat tokens, for which the preferred responses were bert and bäht. This finding is also somewhat unexpected since production and perception studies have reported that German /e/ is substituted for English /æ/ (Arndt & Careless, 1978; Oakeshott-Taylor, 1976; Schröder, 1979; Weiher, 1975; Wode, 1978, 1981). This would lead one to expect that bet would be the preferred response. The fact that bert and bäht, whose vowels are longer (and lower) than the one in bet, were given as responses is interesting because it shows that L1 German listeners can perceive a difference between the vowels in bet and bat. Although the N response was rarely chosen (which would have indicated that none of the supplied alternatives was appropriate), the low confidence ratings for bat indicate that none of the German words were judged to have a vowel that was a close counterpart for the vowel in bat.

The perceptual result for bat is difficult to interpret in the light of the acoustic comparison of the English stimuli to the German subjects' productions of the response categories. Figure 1 shows that the English [æ] tokens are within the area covered by the vowel in bäht and at the boundary of the ellipsis for the vowel in bert. One possible explanation for the low confidence ratings for bat is that the English [æ] tokens are located at or close to the periphery of two vowel areas that are quite unstable in the dialect of the listeners, as shown by the large standard deviations in Table 2 and the large areas in Figure 1 for the vowels in bäht and bert. As pointed out earlier, the relatively low [e:] pronunciations for (ä) obtained in this study are atypical
for northern Germans. The highly variable pronunciations for (er) across subjects probably reflect different social evaluations of the low (er) pronunciations that are stereotypically associated with a Kiel accent. It seems likely, therefore, that the vowel in *bat* was not judged to have a close counterpart in German because of the unstable character of the German subjects' vowel categories in the acoustic vicinity of English [æ].

IDENTIFICATION EXPERIMENT

Several hypotheses about the perception of English /i, ɪ, ɛ, æ/ by L1 German speakers can be proposed based on the findings presented thus far. A hypothesis derived from Flege's model of L2 speech learning (Flege, 1987, 1988; Flege & Hillenbrand, 1984) is that since English /i/ and /ɪ/ are highly similar (or identical) to their German counterparts, native German listeners should perceive the English /i-/ɪ/ contrast much like native English listeners. If they differed slightly, their perception of these English vowels should not be affected much by extended exposure to English. The English /ɛ/-/æ/ contrast, on the other hand, involves the new vowel /æ/. Inexperienced Germans may differ from native English speakers in perceiving this contrast. However, if Flege's hypothesis that adult L2 learners establish phonetic categories for new sounds in an L2 is correct, then extended exposure to English should lead native German speakers to perceive this contrast in an English-like manner. The experiment presented here tested these predictions in a listening experiment in which two groups of L1 Germans differing in experience with English identified members from natural-sounding synthetic *beat–bit* and *bet–bat* continua.

METHOD

Stimuli

Two continua, one ranging from *beat* to *bit* and the other from *bet* to *bat*, were created using the parallel mode of the Klatt (1980) software synthesizer. Two parameters were varied factorially: duration in 3 linearly equal steps and vowel spectrum in 11 linearly equal steps, so that 33 stimuli each were generated for the *beat–bit* and *bet–bat* continua.

For the *beat–bit* continuum, the vowel portion varied from values appropriate for English /i/ to those for /ɪ/. The nominal values for formants one to three (F₁-F₃) in the endpoint /i/ stimulus were: 310 Hz, 2,020 Hz, and 2,960 Hz. The /ɪ/ endpoint had nominal F₁-F₃ values of 400 Hz, 1,900 Hz, and 2,570 Hz. F₄ and F₅ were set at 3,300 and 3,850 Hz. The bandwidths for the five formants were: 90, 110, 170, 250, and 200 Hz. F₀ remained at 120 Hz for the first half of the vowel portion, then was interpolated to 100 Hz over the final half of the vowel.

The vowel portion of each of the 11 spectrally different stimuli had nominal durations of 150, 200, and 250 ms. (The actual durations as measured with a waveform editor were 138, 190, and 233 ms.) These durations included the formant transitions used to cue the word-initial and word-final stops.
The initial /b/ was synthesized by interpolating $F_1$ from 180 Hz, $F_2$ from 1,465 Hz, and $F_3$ from 2,180 Hz to the steady-state values for the following vowel. The $F_1$-$F_3$ transitions were 25, 45, and 50 ms in duration. The final /t/ was simulated by interpolating from the vowel steady-state to 300 Hz for $F_1$, 2,000 Hz for $F_2$, and 2,900 Hz for $F_3$. The durations of the word-final $F_1$-$F_3$ transitions were 40 ms, 40 ms, and 60 ms. A 65-ms period of silence occurred between the end of the /t/ transitions and a wide-spectrum 30-ms release burst. Even though /t/ is not always released in the final position of English words, the release burst was added because a clear /t/ percept could not be obtained without it.

The bet-bat continuum was identical to the beat-bit continuum except for the $F_1$-$F_3$ vowel portions, which varied from values appropriate for English /æ/ to those for English /œ/. The nominal values for $F_1$-$F_3$ in the endpoint /æ/ stimulus were: 515 Hz, 1,830 Hz, and 2,490 Hz. The /œ/ endpoint had nominal $F_1$-$F_3$ values of 675 Hz, 1,730 Hz, and 2,400 Hz (see note 4).

The peak syllable intensity of the 66 stimuli (2 Continua $\times$ 11 Spectral Steps $\times$ 3 Durations) was normalized using a waveform editor.

Subjects

Three groups of 10 listeners each (5 males, 5 females) participated as paid subjects. Subjects in the “inexperienced” native German group (GA) had been in an English-speaking environment for a mean of 0.6 years, whereas the mean for the “experienced” native German group (GB) was 7.5 years. The mean age of the GA and GB subjects was 28 and 33 years, respectively. Subjects in GA and GB had studied English in school for about the same number of years (7.6 and 6.6 years). One GA and one GB subject reported that they were not currently using German, but all others used German daily at the time of the study. The mean self-estimated percentage of daily use of English was 66% for GA and 87% for GB. The native English control group (EN) consisted of monolingual subjects (mean age of 28 years). All subjects passed a standardized hearing test prior to the experiment.

Because it has been suggested that listeners from southern Germany make greater use of duration in vowel perception than northern Germans, it would have been desirable to have a dialectally homogeneous German subject population. However, due to limited German subject availability in the Birmingham, Alabama, region, the German subjects in the present study came from various German areas. Judging from their origin, they represented the following broad dialectal backgrounds, namely, northern German (GA: 3, GB: 3), central German (GA: 5, GB: 4), and southern German (GA: 2, GB: 3).

Procedure

The digitized waveforms were low-pass filtered at 4.8 kHz. Eleven randomizations of the stimuli in each continuum were presented binaurally (TDH-49) at a comfortable listening level (72 dB) in a self-paced experiment. The order of presentation of the two continua was counterbalanced. The sub-
Subjects identified the stimuli as *beat* or *bit*, or as *bet* or *bat*, by pressing one of two buttons of a response box connected to a PDP11/73 computer. Average running time for the experiment was 30 min.

The frequency with which subjects responded *beat* (to the *beat*–*bit* continuum) and *bet* (to the *bet*–*bat* continuum) was tabulated. Responses to the first of the eleven randomizations were not analyzed. The resulting 33 scores for the two continua were submitted to separate Group × Duration × Spectral Step (3 × 3 × 11 levels) ANOVAs. Further analyses were carried out for subjects showing a crossover as a result of spectral changes. A crossover was said to occur if the differences between the responses to the endpoint stimuli (1 and 2 versus 10 and 11) was 70% or greater. This criterion was applied separately to stimuli with short, medium, and long vowel durations.

Responses that met this criterion were subjected to a probit analysis,\(^5\) which provided a phoneme boundary score and a slope value. The phoneme boundary score indicates the point at which the continuum is separated into two perceptual categories. The slope value is a measure of the steepness of the identification function. Large values are obtained for steep functions indicating certainty in the identification of stimuli. Conversely, small values are obtained for shallow slopes, which usually indicate uncertainty for stimuli in the region of the phoneme boundary. The phoneme boundary scores and slope values were submitted to Group × Duration ANOVAs testing the effect of duration on the location of the phoneme boundary and the steepness of the identification functions.

**Results**

**Beat–bit continuum.** Figure 2 presents the group identification functions for the *beat*–*bit* continuum for EN, GB, and GA, averaged over the three vowel durations. The three listener groups identified the endpoints predominantly as *beat* and *bit*, respectively, suggesting that the spectral differences in the stimuli were sufficient to cue the /iː–ɪ/ contrast for all three groups. However, as suggested by the large standard deviations for GB and GA in Figure 2, the identification functions of individual German subjects differed considerably.

The differences among the German subjects were probably not related to their regional origin. The responses of 6 (of 20) subjects from northern Germany were compared to those of 5 subjects from southern Germany. (Subjects from central Germany were ignored in order to compare regional extremes.) A Region × Duration × Spectrum (2 × 3 × 11 levels) ANOVA revealed no main effect or interaction involving Region \( (p > .1) \). This indicates that the 6 northern and 5 southern Germans made much the same use of vowel quality and duration in identifying English /iː/ and /ɪ/. Although it has previously been reported (Weiss, 1972, 1976) that regional differences exist in native Germans' use of vowel quality and duration to differentiate pairs like /iː–ɪ/, it appears doubtful that a larger population might have revealed differences since the Region factor and all interactions involving

\(^5\) Probit analysis is a statistical method for analyzing dichotomous outcomes.
Figure 2. Group identification functions for the beat-bit continuum, averaged over three vowel durations. Error bars indicate ± 1 standard deviation.
Region missed significance by a large margin. Figure 3 shows the results from 6 GA subjects, arranged according to their regional origin (top panels: northern Germany; middle panels: central Germany; bottom panels: southern Germany). The subjects on the left side of Figure 3 showed fairly sharp crossovers from /i/ to /ɪ/ responses as spectral quality changed, whereas the subjects on the right side of Figure 3 showed little or no systematic effect of spectral manipulation.

Figure 4 shows that all 10 native English subjects and 11 of the 20 German subjects were clearly more influenced by spectral than by duration differences. Their datapoints cluster in the upper left corner of Figure 4 because they showed less than a 25% change in responses due to duration, and more than a 60% change due to spectral differences. The datapoints of 4 GB and 5 GA subjects appear on the right side of Figure 4 because they showed at least 50% change in labeling as a result of the duration manipulation and in most instances less than a 40% change as a result of the spectral manipulation.

A three-way ANOVA yielded significant interactions of Group × Duration, $F(4, 54) = 2.999$, $p < .05$, and Group × Spectrum, $F(20, 270) = 1.789$, $p < .05$. The Group × Duration interaction was obtained because the change in beat responses as vowel duration increased from 150 to 200 to 250 ms was smaller for EN (9%) than for the German groups (GB: 31%; GA: 35%). Tests of simple main effects revealed that all three groups showed significant response changes due to the duration manipulation ($p < .05$). Post-hoc tests for each group revealed that the three groups responded less often with beat for stimuli with short vowels (EN: 42%; GB: 39%; GA: 29%) than for those with medium (EN: 47%; GB: 65%; GA: 57%) or long duration (EN: 51%; GB: 70%; GA: 64%) (Newman-Keuls, $p < .05$). The Group × Spectrum interaction was obtained because the difference in beat responses between the spectral endpoint stimuli was larger for EN (88%) than for GB and GA (64%, 63%). Tests of simple main effects revealed that the response changes due to the spectral manipulation were equally significant for all three groups ($p < .001$).

These findings suggest that the perception of the English /i/-/ɪ/ contrast is based primarily on spectral quality for EN, whereas GB and GA differentiate this contrast on the basis of both spectrum and duration. No differences were observed between GB and GA in the relative use of spectral or temporal cues, although these groups differed in amount of English-language experience. To further explore between-group differences, data from subjects who showed a crossover as a result of spectral changes were further analyzed.

The total number of spectrally based crossovers in beat versus bit judgments for stimuli with short, medium, or long vowel duration was 18 for GB and 16 for GA, as compared to 29 for EN (maximum = (3 vowel durations x 10 subjects) = 30). The small difference among GB and GA and the relatively large difference between the two German groups and EN indicates that prolonged exposure to English does not increase German listeners’ use of spectral cues to differentiate the English /i/-/ɪ/ contrast.

The phoneme boundary scores and slope values computed for identifica-
Figure 3. Examples of individual subjects' identification functions for the *beat-bit* continuum, arranged according to the steepness of the identification function (left vs. right) and regional origin (top: northern Germany; middle: central Germany; bottom: southern Germany).
Figure 4. Effect of duration and spectral differences on change in identification for the beat-bit continuum. EN = native English subjects; GB = experienced native German subjects; GA = inexperienced native German subjects.

tion functions that showed crossovers were submitted to Group (German vs. English) × Duration (3 levels) ANOVAs. No interaction or main effect at \( p < .05 \) was obtained for the slope values, which indicates that those German subjects who used spectral cues rather than relying mostly or exclusively on vowel duration were as confident in their use of spectral cues as the EN subjects. The ANOVA for the phoneme boundary scores yielded a significant effect of Duration, \( F(2, 32) = 9.708, p < .001 \), because both groups gave more /i/ responses for stimuli with short vowels than medium or long vowels. On average, the shift from /i/ to /u/ occurred almost one step earlier for the short vowels (at 5.0) than for the medium (5.7) and long (5.9) vowels. Neither Group nor Group × Duration reached significance at \( p < .05 \) for the phoneme boundaries, indicating that the German subjects who showed a crossover divided the continuum at about the same point as the English subjects.

Bet-bat continuum. Figure 5 presents the group identification functions for the bet-bat continuum for EN, GB, and GA, averaged over the three vowel durations. The endpoint stimuli were unambiguously identified by the EN subjects, whereas subjects in both GB and GA were quite variable as suggested by the large standard deviations.

Figure 6 provides individual subject information about the relative importance of spectral and temporal cues to the /e/-/æ/ contrast. All 10 EN subjects showed a larger shift in labeling as the result of spectral rather than duration differences, whereas the datapoints of GB and GA form a continuum from predominantly spectral to predominantly temporal effects on the identification of stimuli as bet or bat. The datapoints from GB and GA suggest that duration becomes a less important cue as spectral information becomes more important. Figure 6 also reveals that more datapoints from
Figure 5. Group identification functions for the bet-bat continuum, averaged over three vowel durations. Error bars represent ± 1 standard deviation.
GB than GA are located near the cluster formed by EN, indicating that GB listeners were affected less by duration and more by spectral differences than GA listeners.

A three-way ANOVA yielded a significant interaction of Group × Duration × Spectrum, $F(40, 540) = 2.223, p < .001$, in addition to significant Group × Duration, $F(4, 54) = 7.300, p < .001$, and Group × Spectrum, $F(20, 270) = 11.487, p < .001$, interactions. The three-way interaction was explored through separate ANOVAs for each duration (short, medium, long), testing the effect of Group on the change in bet responses between the spectral endpoint stimuli. The response changes due to the spectral manipulation were significant for each group in all three durations ($p < .01$). For stimuli with short vowel duration, the change in labeling between the spectral endpoint stimuli was smaller for GA (41%) than EN (97%), whereas the change for GB (67%) did not differ significantly from either GA or EN (Newman-Keuls, $p < .05$). For stimuli with medium vowel duration, the effect of the spectral manipulation was smaller for GA (37%) than for EN (98%) and GB (76%), who did not differ significantly from each other (Newman-Keuls, $p < .05$). For stimuli with long vowel duration, the change in bet responses as a result of the spectral manipulation was larger for EN (99%) than for both GB (67%) and GA (31%), who did not differ significantly from each other (Newman-Keuls, $p < .05$).

These results show that the effect of spectral manipulation was significantly smaller for GA than EN for all three durations, whereas this effect did not differ significantly for EN and GB for two durations (short and medium). The conclusion that prolonged exposure to English leads native Germans to perceive the English /e/-/æ/ contrast in a more native-like way receives further support from the exploration of the Group × Duration interaction. The response changes resulting from the duration manipula-
tion, which were significant for all three groups ($p < .001$), were smaller for EN (15%) than for GB (45%), which were, in turn, smaller than for GA (59%).

These findings indicate that the three groups differed with respect to the relative effect of spectral and duration manipulations in judging vowels as /e/ or /æ/. The spectral cues were more important for EN than for GB, and more important for GB than for GA. Conversely, the duration cue was more important for GA than for GB, and more important for GB than for EN. These results suggest that English language experience affects how native German listeners perceive English /e/ and /æ/. To further explore between-group differences, data from subjects who showed a crossover as a result of spectral changes were further analyzed.

The total number of spectrally based crossovers in bet versus bat judgments for stimuli of short, medium, or long vowel duration was 18 for GB and 6 for GA, as compared to 30 for EN (maximum = (3 vowel durations x 10 subjects) = 30). The fact that more crossovers were observed for GB than GA again indicates that English language experience may cause native German listeners’ perception of /e/ and /æ/ to become more English-like.

Phoneme boundary scores and slope values for identification functions that showed crossovers were submitted to Group (German vs. English) x Duration ANOVAs. A significant Group x Duration interaction, $F(2, 30) = 6.440, p < .01$, was obtained for the phoneme boundary scores because the German subjects’ crossovers for stimuli with medium vowel duration occurred one spectral step closer to /e/ (at 5.3) than the crossover for EN (at 6.3) ($p < .05$), and because the German listeners’ crossovers for the long vowels occurred 1.4 spectral step closer to /e/ (at 4.2) than the crossover for EN (at 5.6) ($p < .001$). The crossover for the short vowels was the same for both groups (at 7.2). This means that the native Germans gave more /æ/ responses than the native English listeners for stimuli with medium and long vowel durations.

The ANOVA examining slope values yielded no interaction or main effects at $p < .05$, indicating that those German subjects who used spectral cues rather than relying mostly or exclusively on vowel duration were as confident in their use of spectral cues as the EN subjects.

GENERAL DISCUSSION

The most important result of the present study was the finding that L2 experience influenced the perception of two L2 vowel contrasts differently, and that these differences could be accounted for in terms of cross-language sound correspondences. Two groups of native German listeners differing primarily in English-language experience showed no differences in the identification of a beat–bit continuum, whereas clear differences between these groups were observed for a bet–bat continuum.

An approximately equal number of experienced and inexperienced native German listeners based their identification of stimuli from the beat–bit
continuum mostly on duration differences. Those native German listeners who identified the beat-bit continuum primarily on the basis of spectral differences, as did all members of the native English control group, came from both German groups. This suggests that extended contact with English as a second language does not affect the perception of English /i/ and /ɪ/ by native Germans.

The results for the bet-bat continuum, on the other hand, showed that a larger number of experienced than inexperienced German subjects identified stimuli in a way that was similar to the native English listeners. That is, more experienced than inexperienced German subjects relied on spectral rather than temporal cues, suggesting that extended contact with English may precipitate the English-like perception of the /e/-/æ/- contrast.

The difference between experienced and inexperienced German listeners provides insight into the nature of adult speech learning. One general claim about the effect of L2 experience on adult speech learning derives from the critical period hypothesis which, among other things, predicts that extended exposure to an L2 cannot overcome maturationally conditioned limits on the ability of adults to learn (an) additional language(s) successfully.

Support for this claim, as far as it relates to L2 vowel perception, would seem to come from a study by Sitaras and Gottfried (1984), who found that differences in L2 French experience of their L1 English subjects were not reflected in differences in the identification and discrimination of L2 vowel contrasts. However, the results of the second part of the present study show that this is not true of all vowel contrasts. At least some of the native Germans had learned to perceive the English /e/-/æ/- contrast in a native-like way. Importantly, differences among the native Germans were clearly related to amount of L2 experience. The seemingly contradictory results of the present study and that of Sitaras and Gottfried (1984) could be due to the fact that the experienced and inexperienced groups in this study differed greatly in exposure to English (0.6 vs. 7.5 years in the United States), whereas subjects in the Sitaras and Gottfried study differed in terms of level of foreign language instruction. It seems likely that several years of L2 input from native speakers, rather than a number of foreign language classes, are necessary to change adult learners’ L2 vowel perception.

Results from the identification of the beat-bit continuum in the second part of the present study were consistent with the hypothesis that the perception of a vowel contrast involving similar vowels (/i/, /ɪ/) would not be affected by amount of L2 experience. For the native English listeners, the /i/-/ɪ/- contrast was primarily spectrally defined, which is consistent with previous findings (Stevens, 1959a, 1959b). Some native Germans, both inexperienced and experienced, responded like the native English listeners, while others were strongly influenced by the duration cues. Individual differences in the use of duration versus spectral cues were also reported by Weiss (1972, 1976) for native German listeners’ perception of the German /i/-/ɪ/- contrast.

The results from the native Germans for the beat-bit continuum, which are reflected in an apparently bimodal distribution in Figure 4, lead to the
following question: Is it the case that some Germans learn the native English-like perception of the /i/-/ɪ/ pair almost instantaneously upon contact with English, while others do not learn it at all, irrespective of L2 experience? Or, do native Germans simply use the same perceptual cues to differentiate the /i/-/ɪ/ contrast in English as they do in their native language? Perhaps those Germans who happen to differentiate German /i/-/ɪ/ on the basis of spectrum get a “free ride” and do not have to learn anything with respect to the English /i/-/ɪ/ contrast, while those who differentiate German /i/-/ɪ/ on the basis of duration do not modify their primarily temporally defined prototypes.

The latter alternative seems to be the more plausible one. It receives some support from the experiment in the first part of the present study, where it was shown that native German listeners identified English *beat* and *bit* tokens highly consistently and confidently with the German words *biet* and *bitt*, respectively, even though the vowel durations in *beat* and *bit* overlapped. This suggests that for native German listeners, spectral information is sufficient to differentiate the /i/-/ɪ/ contrast in the absence of clear temporal cues. It seems unlikely, therefore, that the ability to differentiate English /i/ and /ɪ/ using spectral cues is a consequence of L2 contact. We can probably assume that German learners of English do not have to acquire the use of spectral cues to differentiate English /i/ and /ɪ/. The duration difference between German /i/ and /ɪ/, which in some German dialects may already be small in stressed syllables (Barry 1974a, 1974b), tends to be even smaller, if not nonexistent in unstressed position (Wängler, 1967; Zwirner, 1959). Unless the paradoxical assumption is made that those Germans who relied on the duration cue in the identification of the *beat–bit* continuum cannot make use of a cue (spectral information) in English that they must use in perceiving unstressed German vowels, it may be concluded that the predominant use of duration by some Germans to differentiate English /i/-/ɪ/ simply reflected a preferred perceptual strategy for stressed vowels that seems to be immune to L2 experience. This needs to be tested in future research.

Results from the identification of the *bet–bat* continuum were consistent with the hypothesis that L2 experience would affect the perception of a contrast involving one similar and one new vowel (/e/, /æ/). It was found that the experienced German listeners identified the continuum in a more English-like way than the inexperienced German listeners. However, not all of the experienced Germans identified stimuli primarily on the basis of spectral changes as did the native English subjects. In addition, a few inexperienced Germans were unlike the majority of subjects in that group in that they resembled the native English and the majority of experienced German subjects in identifying stimuli more on the basis of spectrum than duration. This indicates that although experience is an important factor in learning to differentiate a new vowel contrast, it is not the only one. Future research will have to determine why some individuals need only a small amount of L2 experience to perceive a new vowel contrast similarly to native speakers of the L2, while others seem to be virtually immune to L2 experience (see also Underbakke, Polka, Gottfried, & Strange, 1988).
The experiment in the first part suggested that native Germans with relatively little exposure to English perceive a difference between English /e/ and /æ/. These subjects identified *bet* mostly with *bett* and *bat* with *bert* or *bäht*. The experiment in the second part further indicated that the primary perceptual cue to the /e/-/æ/ contrast for inexperienced native Germans is duration. The results of some cross-language studies of vowel perception have been interpreted to mean that reliance on the duration cue depends on the phonemic use of duration in the listeners’ native language (Barry, 1974a, 1974b; Bennett, 1968; Gottfried & Beddor, 1984, 1988; Janson, 1979; Lamminmäki, 1979; Stevens, Liberman, Studdert-Kennedy, & Öhman; 1969). The use of duration by the inexperienced Germans in the second part could therefore be viewed as a German perceptual strategy that is transferred to L2 English. Studies of German vowel perception by native German listeners have shown that vowels from the upper portion of the vowel area are differentiated primarily on the basis of spectral cues, whereas nonback low vowels are identified primarily on the basis of duration. Thus it would seem that the inexperienced native Germans’ overuse of duration to differentiate English /e/-/æ/ may have been due to transfer of an L1 perceptual strategy.

However, there is an alternative explanation for the use of duration by the inexperienced subjects that does not invoke L1 transfer. Vowel perception studies have shown that if qualitative differences do not provide listeners with clear perceptual cues, they will differentiate vowels on the basis of duration (Bennett, 1968; Lieberman & Kubaska, 1979; Weiss 1976). In a study of the perception of Cardinal Vowels by native speakers of British English and German, Butcher (1976) reported that the perceived distance among low front vowels was larger for English than for German subjects. He accounted for this by pointing out that English vowels are acoustically close together in the low front portion of the vowel space, whereas German vowels in that area are widely separated. Butcher’s study, which also compared different age groups of native English and German listeners, suggests that the arrangement of German vowels in the acoustic vowel space “desensitizes” native German speakers to spectral differences among low front vowels. Since native Germans lack L1 experience with these vowels, it is to be expected that qualitative differences among low front L2 vowels do not provide (inexperienced) native German listeners with clear perceptual cues, so that, in accordance with Bennett’s and Weiss’ findings, duration will take over as a cue.

Instead of attributing the inexperienced native Germans’ reliance on duration in differentiating the /e/-/æ/ contrast to a German perceptual strategy, their response pattern might be accounted for by a language-independent perceptual principle. This principle states that whenever spectral differences do not provide sufficient cues to differentiate vowel contrasts for whatever reason (e.g., linguistic desensitization), duration differences are relied upon. This hypothesis is consistent with the findings of Bennett (1968), Weiss (1976), and the present results seen in the light of Butcher’s (1976) findings.

This hypothesis can also be used to account for the results of a study by Gottfried and Beddor (1984, 1988), which led the authors to suggest that the
use of the duration cue depends on the phonemic status of duration in the listener's L1. Gottfried and Beddor reported that native (American) English, but not native French, listeners used duration to differentiate the French /ɔ/–/o/ contrast. They attributed their finding to the fact that duration is a more reliable cue to English than French vowel contrasts. However, in a comparison of adult native French and (British) English listeners' perception of Cardinal Vowels, Butcher (1976) found that for the English listeners, back vowels were perceptually close together, whereas native French listeners perceived them to be much further apart. It can be assumed that the perceptual distance between back vowels would be even smaller for speakers of American English because most speakers of this dialect group lack an /ɔ/ category. Native speakers of English, especially American English, are thus not as sensitive to spectral differences among back vowels as native French speakers. As predicted by this hypothesis, the native English subjects in the Gottfried and Beddor (1988) study relied more on duration than the native French subjects to differentiate the back vowels /ɔ/ and /o/.

The most convincing argument for this hypothesis and against the view that reliance on the duration cue is an L1 perceptual strategy comes from studies in which L2 learners whose L1 does not have duration contrasts use duration to differentiate vowel contrasts in the L2. Flege and Bohn (1989) reported that native Spanish speakers who identified the same continua as the native Germans in the present study differentiated among a pair of English vowels (/ɪ/-/ɨ/) from an area where Spanish has only one vowel (/i/) primarily on the basis of duration. Since Spanish has no duration contrast, the use of this cue to differentiate a foreign vowel contrast cannot be due to perceptual strategy transferred from L1. The fact that Spanish has only one category in a region of the vowel space where English has two makes it seem likely that, by these arguments, native Spanish listeners are not as sensitive to spectral differences among close front vowels as native English listeners. The reliance on duration to differentiate English /ɪ/-/ɨ/ in the Flege and Bohn (1989) study is consistent with the hypothesis developed earlier.

**CONCLUSION**

The present study showed that L2 experience affects the perception of an L2 contrast that involves a vowel for which no counterpart exists in L1. No such effect was observed for an L2 contrast involving vowels that are similar (or identical) to L1 vowels. This result bears out the predictions of the initially stated hypothesis and is consistent with Flege’s model of L2 speech learning (cf. Flege, 1987, 1988; Flege & Hillenbrand, 1984). It was found that experienced and inexperienced native German learners of English did not differ with regard to the predominant use of either temporal or spectral information to differentiate the English /ɪ/-/ɨ/ contrast.

However, a greater number of experienced than inexperienced native German listeners perceived the English /ɛ/-/æ/ contrast like the native English control subjects, who relied primarily on spectral cues. Differing from previous assumptions about reasons for the overuse of duration in differentiat-
ing an L2 vowel contrast, it was hypothesized that the use of the duration cue does not indicate reliance on an L1 perceptual strategy but rather reflects a general speech perception strategy that takes over whenever information conveyed by spectral differences is insufficient.

ACKNOWLEDGEMENTS
This study was supported by a grant from the National Institutes of Health (NS20963-04). Part of the results reported were presented at the annual meeting of the Linguistic Society of America, New Orleans, LA, December 27-30, 1988. The authors wish to thank W. Barry, E. Fischer-Jorgensen, and W. Moulton for their criticism and helpful comments.

NOTES
1. The term “English” as used in this article refers to American English unless stated otherwise.
2. It should be pointed out, however, that /æ/ is part of the phonemic and phonetic repertoire of several German dialects, e.g., Bavarian (Traummüller, 1982), Swiss German (Boesch, 1957), and dialects spoken along the central Rhine valley (Pilch, 1966). An [æ(:)] may also be observed in North German dialects in which vocalization of /r/ may result in a lowering and lengthening of the preceding vowel vis-à-vis its /r/-less counterpart. For example, the vowel in bert as pronounced by 5 (of 11) monolingual speakers from the Kiel (northern Germany) area was transcribed as [æ] or [æ:] by four phonetically trained American English listeners.
3. The decision to present biet, bitt, bett, bāht, bert, and N as response alternatives resulted from a pilot study that indicated that these responses, but no others (e.g., bart), would be used. The alternatives are not all real words in German, but their pronunciation is fairly unambiguous because of the phonemic nature of German orthography. The vowels in biet, bitt, bett are /i/, /ɪ/, and /ɛ/, respectively. Orthographic (āh) is usually /e:/ in northern German. However, the vowel in bāht may be /ɛ:/ even for speakers from northern Germany (who do not normally use it; cf. Kohler, 1977; Moulton, 1962, 1987; but may have it as a perceptual category according to Sendmeier, 1981), probably because this form is derived from an onomatopoeic expression indicating the sound sheep make. Bert was included as a response category because listeners from the Kiel area frequently hear mid-low to low [ɛ:]≈[æ(:)] pronunciations for orthographic (er). Pronunciations of (er) may vary considerably across speakers and styles in northern Germany. Apart from the relatively formal and socially more prestigious [ɛ:]≈[æ(:)] pronunciations of (er), there are monophthongal [ɛ:]≈[æ(:)] pronunciations that are associated with an informal style of speaking, a strong regional (Kiel) accent, and perhaps low social status.
4. Peterson and Barney (1952) male values were used for stimuli 2 and 10. The values for the endpoint stimuli 1 and 11 were extrapolated. The nominal synthesis values were confirmed through LPC analysis.
5. Probit analysis (Bock & Jones, 1968; Finney, 1971), a maximum likelihood technique, fits cumulative normal distribution functions to the data and provides estimates of the 50% crossover point (i.e., the phoneme boundary) and of
the standard deviation of the crossover point, which is inversely related to the steepness of the cumulative curve.

REFERENCES


(forthcoming). The production of new and similar vowels by adult German learners of English.


Sajavaara (Eds.), *Papers in contrastive phonetics* (pp. 165–231, Jyväskylä Cross-Language Studies 7).


