The detection of French accent by American listeners

James Emil Flege
University of Alabama, Birmingham, Alabama 35294

(Received 19 January 1983; accepted for publication 22 March 1984)

The five experiments presented here examine the ability of listeners to detect a foreign accent. Computer editing techniques were used to isolate progressively shorter excerpts of the English spoken by native speakers of American English and French. Native English-speaking listeners judged the speech samples in one- and two-interval forced-choice tests. They were able to detect foreign accent equally well when presented with speech edited from phrases read in isolation and produced in a spontaneous story. The listeners accurately identified the French talkers (63%-95% of the time) no matter how short were the speech samples presented: entire phrases (e.g., "two little dogs"), syllables (/tu/ or /ti/), portions of syllables corresponding to the phonetic segments /t/, /i/, /u/, and even just the first 30 ms of "two" (roughly, the release burst of /t/). Both phonetically trained listeners familiar with French-accented English and unsophisticated listeners were able to accurately detect accent. These results suggest that listeners develop very detailed phonetic category prototypes against which to evaluate speech sounds occurring in their native language.

PACS numbers: 43.70.Dn, 43.70.Ve

INTRODUCTION

Humans are able to detect subtle variations in the vocal signals used for communication. From a purely analytical perspective we can identify several different types of information present in human speech. In addition to acoustic information used for the phonetic categorization of sounds, humans also derive information concerning speaker-related properties such as personal identity, age, gender, attitudes, mood, and state of health. For example, listeners can identify gender at better than chance levels based on children's production of isolated vowels (Sachs, 1975). Listeners are also able to determine whether a talker differs from themselves in social or geographic origin (see Trudgill, 1975).

Unlike most previous speech perception research, this study focuses on listeners' ability to detect differences between groups of talkers differing in language background rather than on the identification of phonetic categories. Specifically, it examines the ability of American listeners to determine whether short samples of English spoken by native and French talkers were produced by a non-native speaker of English.

There are two reasons to study accent detection. As synthesis-by-rule programs become more sophisticated, development efforts will need to concentrate on producing speech that is natural sounding and easily comprehensible. This requires an understanding of what constitutes the perceptual phonetic norms for the speech sounds being synthesized, and to what extent departures from those norms are acceptable. At a more theoretical level, an understanding of speech perception must include an understanding of how all information relevant to the message—including foreign accent—is extracted from the speech signal.

Most researchers accept that talkers develop central phonetic representations for phonetic categories, and that these representations include an important auditory-based component (MacNeilage, 1980). Nootboom (1973) conceptualizes phonetic representations as "mental images" representing all language-specific aspects of "inner speech." Such representations are thought to include information pertinent to the prosodic, intonational, and coarticulatory characteristics of particular languages, as well as details concerning the norms for segmental timing and place/manner of articulation. It is likely that speakers of different languages, who often produce cognate sounds with systematic differences (Ladefoged, 1980, 1983), develop somewhat different phonetic representations which are optimally suited to the way sounds are typically articulated in their native language. It is presumably this which leads to differences in the way speakers of different languages identify and discriminate the members of synthetic speech sound continua (e.g., Abramson and Lisker, 1970; Lisker and Abramson, 1970; MacKain et al., 1981).

The hypothesis underlying the experiments reported here is that the formation of detailed phonetic representations for sounds in their native language enables listeners to detect accent in the speech of non-native speakers. This hypothesis leads to several expectations. First, it should be easier for the listener to comprehend the speech of fellow native speakers than the accented speech of non-native speakers. Second, it should become progressively easier for listeners to comprehend the speech of non-native or speech-handicapped individuals as they gain experience with the patterns of divergences present in those varieties of speech. Third, listeners who are very familiar with a particular variety of speech (e.g., adult monolingual native speakers) should be better at detecting accent (i.e., divergences from the norms of that speech variety) than children or adult non-native speakers who are less familiar with it. There is, in fact, some sup-

At present we do not know how accent detection and phonetic categorization are related to one another. If these two perceptual functions depend on separate processes, the degree to which speech departs from language-specific phonetic norms might be unrelated to intelligibility. However, if accent detection and phonetic categorization are based on a single perceptual process (Ladefoged, 1967), divergences from the segmental phonetic norms of a language (such as those which cue accent) should also lead to a decrement in intelligibility or processing efficiency (Whalen, 1982) and vice versa. Before we begin to test hypotheses such as this, however, we must first gather more detailed information concerning the human ability to detect accent.

Several studies demonstrate listeners' ability to detect accent, and to reliably and validly gauge its magnitude (e.g., Lane, 1963; Asher and Garcia, 1969; Brennan et al., 1975; Ryan et al., 1977; Giles, 1972; Brennan and Brennan, 1981a,b). It should be noted, however, that most previous studies have examined relatively long stretches of speech (phrases, sentences, or paragraph-length passages) that have been read rather than spontaneously spoken. They do not clearly indicate whether differences in segmental articulation between native and non-native speakers are sufficient to cue the detection of accent. Thus listeners' detection of accent in previous studies may have been based solely on supra-segmental differences in timing, stress, and intonation rather than differences confined to the articulation of particular phonetic segments.

This hypothesis is supported by two general considerations. First, linguists (e.g., Bloomfield, 1933; Trubetzkoy, 1939/1969) have claimed that listeners do not readily notice cross-language phonetic differences in segmental articulation that do not importantly affect the categorization of speech sounds in the listener's native language. This suggests that if a foreign language learner produces a sound differently than native speakers, it may not be evident to native-speaking listeners if the sound is perceived as intended. Second, a great deal of speech perception research indicates that speech sounds, especially consonants, are "categorically" perceived (Repp, 1984). This often means that an acoustic difference distinguishing sounds which have been given the same category label will be less readily detected than an acoustic difference distinguishing sounds that have been labeled differently. It suggests that phonetic differences between native and non-native speakers which do not lead to a perceived change in categorical identity (i.e., segmental sound "substitutions") will be poorly detected by native-speaking listeners. If so, English speakers, for example, may have difficulty detecting accent in a Frenchman's production of /\'v/, /\'l/, and /\'u/ when these phonetic segments are articulated in a French-like, albeit recognizable, manner.

The aim of this study was to determine if such subcategorical phonetic differences between native and non-native speakers will suffice to cue the detection of foreign accent. Increasingly short excerpts of the English spoken by native speakers of English and French were presented to native English-speaking listeners. The phonetic segments produced by the non-native speakers were heard as intended in all instances. The task in two experiments was the forced-choice identification of talkers as "native" or "non-native." In three other experiments, speech samples produced by native and non-native talkers were presented in pairs in two-interval forced-choice tasks.

I. GENERAL METHODS

A. Talkers and subjects

Speech samples were obtained from two groups of talkers. American English was represented by eight female monolingual speakers between the ages of 22 and 32 (mean, 26 years). All were students from the central United States (Chicago—4, Denver—2, Rochester—1, Milwaukee—1) who were enrolled in a speech-language pathology program at the time of the study. The second group consisted of eight native French-speaking women aged 28-48 years who had been living in the United States between 6 months and 31 years at the time of the study (mean, 13 years). Of these talkers, two were from Belgium, four from Paris, and two from other parts of France (St. Etienne and Annecy).

The subjects who judged the speech samples were all native speakers of American English who reported their hearing to be normal. Subjects in the five experiments were differentiated according to their experience with French and French-accented English, and according to phonetic sophistication. Each subject filled out a language background questionnaire before participating in the study.

B. Speech materials

To ensure a representative sampling of French-accented English, the following two sets of English phrases were used in counterbalanced order to elicit speech production by the native and non-native talkers in three successively more demanding tasks:

(1) two little boys (1) TV programs
(2) two little girls (2) TV schedules
(3) two little cats (3) TV ratings
(4) two little dogs (4) TV violence
(5) two little birds (5) TV reception
(6) two little mice (6) TV antennas
(7) two little men (7) TV commercials

The first speaking task called for talkers to simply read phrases from a typed list. In the second task, talkers generated original sentences by completing each phrase. In the third speaking task, they again produced each one of the phrases in the utterance-initial position of a complete, original sentence. This time, however, they were required to link their sentences together so that they formed a story. To facilitate this last speaking task, the seven phrases were typed on 3 X 5 in. cards for talkers to arrange on the table before them as they silently rehearsed their story. Talkers were also permitted to include, as needed, sentences not initiated by one of the phrases in their story in addition to sentences initiated by the seven phrases. To ensure that phrases produced in the story
task occurred in utterance-initial position, the talkers were required to pause before producing each sentence (see Flege and Hillenbrand, 1984).

Prior to the experiment, the talkers were told they were participating in a study of "language creativity" in order to divert their attention away from pronunciation. Systematic debriefing later revealed that none of the talkers realized the study focused on pronunciation.

C. Procedure

The speech samples were recorded on high-quality equipment (Sony, model TCD5M) at 1 ips in a sound treated room with an electret condensor microphone (Nakamichi, CM-300) positioned at about 8 in. from the talker's mouth. The speech material was filtered at 7 kHz (Krohn-Hite, model 3202) before being digitized at a sampling rate of 18 kHz with 12-bit amplitude resolution.

A waveform editing program was used to edit out phrases, syllables, and portions of syllables, and to cross splice the phonetic segments /i/, /u/, and /i/. Segmentation was accomplished by visual inspection of digitized waveforms displayed on the screen of a graphics terminal (Tektronix 4010). The edited waveforms were stored on a high-speed mass storage device for later on-line presentation to subjects. At the time of the experiment, the digitized stimuli were converted to analog form and low-pass filtered at 7 kHz before being presented through headphones (TDH-49) to subjects at a comfortable listening level of about 75 dB(A). The presentation of stimuli and data collection were run under the real-time control of a small laboratory computer (PDP 11/34).

Subjects were tested individually while seated in a sound attenuated chamber before a response box with two buttons. Experiments were self-paced, with the minimum between-trial interval set at 1 s. The rms intensity of all stimuli within a block (and across blocks, where necessary) was adjusted to be equal. The order of stimuli within a block was randomized separately for each subject. There was a maximum of four blocks in an experiment, each lasting 8–12 min. Subjects were not told what proportion of talkers were non-native speakers of English.

The number of times each stimulus was identified as having been produced by a non-native talker was tabulated. The scores were submitted to two different types of analysis of variance. In the "talker" analyses the scores for each of the native and non-native talkers were averaged across subjects (i.e., listeners). Speaker group (i.e., French versus American talkers) served as a between-subjects factor. In the "listener" analyses, scores were averaged across the eight French and the eight American talkers. All of the factors examined in these analyses were repeated measures.

II. EXPERIMENT 1

All eight French talkers were judged by the experimenter to speak English with a French accent despite their long residence in an English-speaking country. However, since previous reports suggest that adults can sometimes pronounce foreign language phrases without accent (Neufield, 1980; Williams, 1980), it was first necessary to objectively verify that the English phrases produced by the French talkers were accented.

A. Methods

1. Subjects

The ten listeners in this experiment were eight males and two females with a mean age of 32.3 years (sd = 3.3 years). All had at least some training in phonetics. Three spoke French, but only two reported they "frequently" encountered English spoken with a French accent. Four other subjects indicated they heard French-accented English "from time to time," the remaining four "hardly ever."

2. Speech materials

The stimuli in this experiment were the phrases "two little dogs" and "two little birds" produced in the isolated phrase and story speaking tasks by native and non-native talkers. In editing these stimuli, phrase onset was defined by setting a cursor about 3 ms before the sharp increase in waveform energy signaling the release of /t/ in "two." Segmentation of the phrase-final /s/ (in "dogs" and "birds") for stimuli edited from the story speaking task was based on visual criteria, augmented by auditory judgments when visually defined criteria were insufficient.

The productions of six native and six non-native talkers were arbitrarily chosen to represent English and French-accented English, respectively. The overall duration of all the phrases produced by the native English speakers averaged 960 ms (sd = 119 ms); the phrases produced by the non-native speakers averaged 997 ms (sd = 179 ms).

3. Procedure

The 12 talkers' (2 groups x 6) productions of "two little dogs" and "two little birds" in the isolated phrase speaking task were presented in one block. Productions of the same phrases from the story speaking task were presented in another block. The order of presentation of the two blocks was counterbalanced across subjects. The experiment was self-paced, with a minimum intertrial interval of 1 s.

The subjects' task was to identify each phrase as having been spoken by a native (i.e., American) or non-native (i.e., French) talker. Each block contained ten separate randomizations without replacement of the 24 speech stimuli (2 groups x 6 talkers x 2 phrases). This yielded a total of 100 judgments for each of the phrases produced in the phrase and story tasks by each talker; and a total of 2400 judgments for phrases produced by both the native and non-native speaker groups.

B. Results and discussion

The results presented in Table I clearly demonstrate that the subjects were able to determine whether a short phrase had been produced by a native or non-native talker. About 89% of the phrases produced by the French native speakers were correctly identified as "non-native," as against less than 1% of the phrases produced by the native speakers of English.

The phrases produced by the French talkers were cor-
directly identified as non-native over 90% of the time in all but two instances. The notable exception was the phrase "two little dogs" produced by French talker number 2. This talker arrived in the United States at the age of 9, whereas the other French talkers first arrived in an English-speaking country in early adulthood. Oyama (1982b) found that age of arrival in a country where the second language is spoken significantly affects the magnitude of perceived accent in the speech of foreign language learners. It appears that French talker number 2 was able to approximate the segmental and suprasegmental norms of English speech production sufficiently well to "pass" as a native speaker, at least when no especially troublesome phonetic segment is present. Her production of "two little birds" may have been identified as accented due to the presence of /r/ in that phrase (at least as produced by the French talkers). Unlike the other French talkers, who seemed to realize /r/ as the uvular fricative found in French, French talker number 2 appeared to "delete" the /r/ in "birds" in both speaking tasks.

Previous research indicates that speaking task may influence the authenticity of foreign language speech production by non-native speakers (Dickerson, 1975; Beebe, 1980; Oyama, 1982b) just as it affects the speech production of native speakers. For example, formant differences distinguishing native-produced vowels are reduced as the speaking task shifts from the production of isolated words, to words read in a paragraph, to spontaneous speech (Koopmans-van Beinum, 1983). This leads to the expectation that accent will be more evident for speech read in phrases than spoken spontaneously in a story.

There was, however, no difference. About 90% of the French-produced phrases from the phrase task were correctly identified as "non-native," as against 88% of the phrases from the story task. This is consistent with a recent finding by Flege and Hillenbrand (1984) that the French syllables /tu/ and /ty/ read by native English speakers in isolated phrases and spontaneously produced in a story were equally identifiable.

III. EXPERIMENT 2

Experiment 1 showed that listeners could readily detect accent in short phrases produced by non-native speakers. The assumption underlying experiment 2 and those which follow is that accent detection will be more difficult for shorter stretches of speech. The results of experiment 1 suggest that accent was as easily detected in phrases that had been read in isolation by non-native speakers as in phrases produced in a spontaneous story. The lack of an effect of speaking task may simply have been a ceiling effect, since even in short phrases there are potentially many differences in segmental as well as suprasegmental articulation that might cue accent. To assess accent detection in syllable-length stretches of speech, and to further explore the effect of the two speaking tasks, the /tu/ syllables produced by native and non-native speakers in the phrase and story tasks were presented to listeners in a two-interval forced-choice task in experiment 2.

A. Method

1. Subjects

The ten listeners in this experiment were seven males and three females with a mean age of 27.8 years (SD = 4.8 years). As a group, they should be thought of as "sophisticated" since all but one had extensive training in phonetics, linguistics, or French. Four had previously participated in experiment 1. Three spoke French, but only one indicated "frequent" exposure to French-accented English.

2. Speech materials

As illustrated in Fig. 1, /tu/ was edited from the phrases "two little dogs" and "two little birds." The segmentation of /u/ from the following /l/ was based on changes in waveform shape and/or intensity. Visual criteria did not provide an unambiguous basis for segmentation in about 15% of the syllables. A perceptual criterion was used in these instances. The cursor of a graphics terminal was moved leftwards one pitch period at a time until neither an "l" sound nor "l-coloring" was perceptually evident in the periodic portion of the syllable. Edited in this way, the /tu/ syllables produced by the native English speakers were somewhat longer (202 ms, SD = 37 ms) than those produced by the native speakers of French (152 ms, SD = 36 ms).

3. Procedure

Syllables edited from phrases produced in the phrase and story speaking tasks were presented in separate blocks, the order of which was counterbalanced across subjects. On
FIG. 1. Segmentation of the /u/ from /l/ interval in the phrase "two little" produced by four different talkers. The visual criteria were changes in waveform amplitude and shape. Auditory appraisal was also used when necessary, as in waveform (d).

Each trial, one of the /tu/ syllables produced by a native English speaker was paired with a /tu/ syllable produced by a non-native speaker. In each of the two blocks, the 12 /tu/ syllables produced by the French native speakers (6 talker × 2 replicate productions) were paired with each of the 12 /tu/ syllables produced by the native speakers of English, once in the first position of the pair, and once in the second position. This yielded a total of 288 pairs of syllables per block. The interstimulus interval was set at 1 s, the intertrial interval at a minimum of 1 s.

The listeners were informed that the syllables they would hear had been edited from "two little birds" and "two little dogs," and that one of the two syllables per trial had been produced by a native speaker of American English, the other by a non-native speaker. They were instructed to push the left button on a response box if the first member of the pair had been spoken by a "non-native" speaker, the right button if it was the second syllable.

Subjects' responses were submitted to two analyses of variance. In the "talker" analysis the responses of the ten subjects were combined. Speaker group served as a between-subjects factor, and condition (isolated phrase versus story) and token (phrase #4 or #5) were repeated measures. In the "listener" analysis, the responses of the six talkers per group were combined. The dependent variable was the frequency with which each subject identified phrases produced by the American and French talkers in the phrase and story conditions as non-native. Condition (isolated phrase versus story) and speaker group (American versus French) were repeated measures.

B. Results and discussion

The results presented in Table II demonstrate that accent can be accurately detected in CV syllables produced by non-native talkers. Overall, /tu/ syllables produced by the French native speakers were correctly chosen as the "non-native" member of stimulus pairs 95% of the time. Syllables produced by the Americans were incorrectly chosen only 5% of the time. The effect of speaker group was highly significant in both the "talker" analysis [F(1, 10) = 1982] and the "listener" analysis [F(1, 9) = 4703, p < 0.0001].

Accent was equally detectable in syllables excerpted from isolated phrases and spontaneous stories. The listeners correctly chose the /tu/ produced by non-native speakers in the phrase reading task about 94% of the time, as against about 95% of the time for syllables produced in the story task. Both the "talker" and "listener" analyses indicated that the effect of speaking condition was not significant (p < 0.01).

There was little difference between individual talkers. Syllables produced by the American talkers were incorrectly chosen as "non-native" 3%-8% of the time. Syllables produced by the French native speakers were correctly chosen as "non-native" 86%-98% of the time. It is interesting to note that /tu/ syllables produced by French talker number 2 were correctly chosen as "non-native" 95% of the time. Recall that, in experiment 1, French talker number 2 was incorrectly identified as "native" 92% of the time when /tu/ occurred in "two little dogs." It is apparent from the results of experiment 2 that her production of /tu/ departed suffi-
ciently from the phonetic norms of English to permit accent detection. The accentedness of her /tu/ may not have been detected in experiment 1 because: (1) it occurred in a longer stretch of speech, (2) the subjects in experiment 1 were less sophisticated phonetically than those in experiment 2, or (3) a paired comparison task represents a more sensitive measure of foreign accent detection than absolute identification.

IV. EXPERIMENT 3

Experiment 2 showed that relatively sophisticated listeners can detect accent in CV syllables produced by non-native speakers. Is this typical of human performance, or were the listeners in experiment 2 able to detect accent so readily because of their special training in linguistics, phonetics, or French? To help answer this question, less sophisticated listeners judged the /tu/ syllables produced by native and non-native talkers in experiment 3. To generalize the findings of experiment 2, the syllable /ti/ was also presented to subjects for accent detection.

Subjects were required to make an absolute identification of talkers as "native" or "non-native." This is probably a less sensitive measure of foreign accent than a paired comparison task because it forces subjects to make use of a less flexible inner criterion. In previous studies untrained listeners have shown a tendency to identify sentences produced by native speakers as having been produced by a non-native speaker (Asher and Garcia, 1969; Scovel, 1981). This may mean that listeners without special training are simply inclined to label speech as "accented" or "distorted" when asked to scrutinize it in an unaccustomed way. However, it may also reflect listener uncertainty concerning what constitutes the phonetic norms of their native language for segmental and/or suprasegmental articulation. It might also indicate that some native speakers depart sufficiently from the norms of their native language to cause subjects to use the label "accent" when it is offered to them in a perceptual experiment. Thus the aim of this experiment was to assess the accuracy with which unsophisticated listeners can detect adherence to, or divergences from, the phonetic norms of their native language in excerpted monosyllables.

A. Method

1. Subjects

The subjects in this experiment had less previous experience that might be expected to contribute to success in an accent detection task than those in experiments 1 and 2. There were two male and ten female undergraduate students with a mean age of 21.6 years (sd = 2.4 years). None had special training in linguistics or phonetics. Although six of the 12 had studied French in high school or college, none could communicate effectively in French or any other foreign language.

2. Speech materials

Since the previous two experiments revealed no difference between stimuli produced in the two speaking tasks, only syllables produced in the phrase task were employed here and in the remaining experiments. In addition to the /tu/ syllables used in experiment 2, this experiment included the /ti/ syllable edited from "TV reception" and "TV antennas." Two additional talkers were also added to the pool of talkers representing American English and French-accented English, making eight per group.

3. Procedure

The task in this experiment was the forced-choice identification of syllables as "native" (i.e., American) or "non-native" (i.e., French). The 16 /tu/ syllables (2 groups X 8 talkers) edited from "two little dogs" and "two little birds" were separately randomized ten times for each subject in two separate blocks, as were the 16 /ti/ syllables. The four blocks were counterbalanced across subjects. The experiment was self-paced, with a minimum intertrial interval of 1 s. The language background of the non-native speakers was not revealed to listeners before the experiment.

The maximum number of choices in the "talker" analysis was 120 (12 subjects X 10 presentations) for each of the 32 /tu/ and 32 /ti/ syllables examined. In the "talker" analysis speaker group (American versus French) was a between-subjects variable and syllable (/tu/ or /ti/) served as a repeated measure. In the "listener" analysis, the maximum number of choices was 160 (8 talkers X 2 replicate productions X 10 presentations). Both speaker group and syllable served as repeated measures.

B. Results and discussion

The results presented in Table III indicate that even unsophisticated subjects can detect accent in English syllables produced by native speakers of French. Overall, syllables produced by the French talkers were correctly identified as "non-native" about 77% of the time. Syllables produced by American talkers were incorrectly identified as "non-native" about 22% of the time. The effect of speaker group was highly significant in both the "listener" [F(1,11) = 410] and the "talker" analyses [F(1,14) = 40.4, p < 0.0001].

In both analyses the interaction of speaker group X syllable reached significance (p < 0.01) because accent was accurately detected more often for /tu/ and /ti/ syllables. The subjects correctly identified French-produced /tu/ syllables as "non-native" 82.4% of the time, and incorrectly labeled /tu/ syllables produced by native speakers as "non-native" 17.1% of the time. For /ti/, on the other hand, they correctly identified the non-native speakers only 70.9% of the time, and misidentified the native speakers as "non-native" 27.5% of the time. However, tests of simple main effects indicated that although more syllables produced by American than French talkers were chosen as non-native, there was no significant difference in the rate of accent detection for /tu/ and /ti/ syllables (p < 0.01).

Most of the syllables produced by individual French talkers were labeled "non-native" far more frequently than those produced by American talkers. The one exception is French talker number 1. Her /tu/ was identified as non-native only 25% of the time. Accent was detected in her /ti/ only 39% of the time, which is much lower than the average correct detection rate of about 85% for the other seven
French talkers. This is somewhat surprising in view of the fact that the phrases from which her /tu/ syllables were edited were correctly identified as "non-native" 98% of the time in experiment 1. This suggests that, for this talker, something in the final three syllables of "two little dogs" and "two little birds" was distinctly non-English, whereas the first syllable (/tu/) came fairly close to English phonetic norms.

Debriefing after the experiment revealed that only one subject was able to correctly identify the language background of the non-native speakers. This, together with the fact that subjects in experiment 3 were not familiar with French or French-accented English, suggests that they detected accent by comparing the speech samples to English phonetic norms, rather than identifying some known characteristics of French-accented English in the speech of the non-native talkers.

V. EXPERIMENT 4

Experiment 3 established that even unsophisticated listeners can distinguish native from non-native talkers on the basis of acoustic differences confined to a single syllable. Experiment 4 used a digital tape-splicing technique to determine whether differences between native and non-native speakers in just one portion of a syllable will suffice to cue accent. Specifically, it examined the ability of subjects to detect accent in just the aperiodic or periodic portion of syllables (/tu/ and /ti/) produced by a non-native speaker.

A. Method

1. Subjects

The listeners in this experiment were ten female students in speech-language pathology with a mean age of 22.5 years (sd = 1.1 years) who were paid for their participation. Nine had taken an introductory course in phonetics, but none spoke French. All ten indicated they generally "pay attention to accents" yet had little exposure to French-accented English.

2. Speech materials

A set of hybrid CV syllables was created by editing out the aperiodic portion (/t/ for short) of the /ti/ and /tu/ syllables produced by eight native and eight non-native talkers in "two little dogs" and "TV antennas." These aperiodic acoustic intervals of variable length were electronically spliced onto the periodic portion (/i//u/) of the syllables /ti/ and /tu/ produced by a female native speaker of English in the same phrases. Thus half of the hybrid CV syllables contained two acoustic intervals produced by a native English speaker, while the other half contained a French-produced /i/ and an American-produced vowel.

Another set of hybrid CV syllables was created by splicing the periodic portion of /ti/ and /tu/ syllables produced by the eight native and eight non-native talkers onto a /t/ edited from either the /ti/ or /tu/ produced by a single native English speaker. The constants /i/, /u/, and /t/ (from /ti/ or /tu/) used to create the hybrid CV syllables were all judged by the author to be representative of Midwestern American English. They were produced under the same con-
FIG. 2. Illustration of the digital cross-splicing technique used to create hybrid syllables in which just one sound (/i/, /u/, or /t/) per syllable was varied. In (a) and (b) a constant /i/ vowel has been spliced onto the /t/ produced by two different talkers; in (c) and (d) a constant /t/ has been spliced onto /i/ vowels produced by two different talkers.

The cross-splicing technique is illustrated for hybrid /ti/ syllables in Fig. 2. The two stimuli at the top represent CV syllables created by appending the periodic portion of /ti/ produced by an American talker onto the aperiodic portions of /ti/ produced by two different talkers. The bottom two stimuli were created by appending the aperiodic portion of /ti/ produced by a single native English speaker onto the periodic portions of /ti/ produced by two different talkers.

The criterion used to segment /t/ from the following /i/ or /u/ was the onset of periodicity. The editing program deleted that portion of the syllable found to the left (or right) of a cursor positioned at the first upward-going zero crossing in the syllable. In splicing together portions of syllables, acoustic segments were rejoined at a zero crossing so that no sudden amplitude change occurred. The hybrid CV syllables created in this way sounded, in the author's estimation, as natural as the unedited CV syllables from which they had been derived.

3. Procedure

Stimuli were presented in the four following blocks, the order of which was counterbalanced across subjects:

Block 1: variable /t/ + constant /i/
Block 2: variable /t/ + constant /u/
Block 3: constant /t/ + variable /i/
Block 4: constant /t/ + variable /u/.

In each block the syllables containing two American-produced phonetic segments were paired with each of the syllables containing one American-produced and one French-produced phonetic segment, once in the first position and once in the second position of the stimulus pair. This yielded a total of 128 trials per block (8 Americans talkers × 8 French talkers × 2 orders), and a total of 160 paired-comparison judgments for each of the 64 hybrid syllables (8 pairings × 2 orders × 10 listeners).

Subjects were not informed that the syllables they would hear represented the combined speech of two different talkers. Their task was to determine which member of the stimulus pair had been produced by a "non-native" (i.e., French) speaker. Subjects indicated their choice by pushing the left or right button on a response box, depending on whether they judged the first or second member of the pair to have been produced by a non-native speaker.

There were three repeated measures in the "listener" analysis of variance: sound type (either the periodic or aperiodic portion of the syllable varied), vowel (/i/ versus /u/), and speaker group (French versus American). The maximum number of choices was 128 (64 French-American pairs × 2 presentations). In the "talker" analysis the maximum number of choices was 160 (8 pairs × 2 repetitions × 10 listeners). Speaker group served as a between-subjects variable; sound type and vowel were repeated measures.

B. Results and discussion

The results presented in Table IV reveal that there is sufficient acoustic information present in a single phonetic segment to permit accent detection. Overall, 67% of the hybrid syllables containing a French-produced segment (/i/, /u/, or /t/) were chosen as the "non-native" member of the pair, as against only 33% of the hybrid syllables containing two American-produced segments. The effect of speaker group was highly significant in both the "talker" [F(1,14) = 117] and the "listener" analysis [F(1,9) = 55.9, p < 0.0001].

It does not appear that accent is detected more readily in vowels than in consonants. The /t/ produced by the French talkers was detected as accented 67% of the time. Accent was also detected 67% of the time in the /i/ and /u/ vowels produced by the French talkers. In neither analysis did the effect of sound type reach significance (p < 0.01).

Consistent with the results of experiment 3, accent was detected somewhat more often in segments edited from /tu/ than /ti/ syllables. In particular, accent was detected more often in French-produced /t/ when this phonetic segment had been edited from /tu/ (71%) than from /ti/ (63%). However, the effect of vowel did not reach significance in either analysis. In the listener analysis the interaction of group × vowel reached significance [F(1,14) = 16.3, p < 0.001]. However, tests of simple main effects indicated that although the effect of speaker group was always significant, the difference between segments edited from /tu/ and /ti/ syllables was not.

These results raise the question of which acoustic characteristic(s) of the stimuli enabled subjects to detect accent. Although it is difficult to directly relate specific acoustic...
attributes of naturally produced speech sounds to perceptual judgments, Table V presents acoustic measurements of /tu/ and /ti/ that are likely to be relevant to the subjects' detection of accent.

The formant center frequencies (F1–F3) of /i/ and /u/ presented in Table V were estimated by LPC analysis, using a 25.6-ms Hamming window whose left margin was positioned at the onset of periodicity of these vowels. The American talkers produced /u/ with significantly higher F2 and F3 values than the native speakers of French, which may be indicative of a more anterior tongue position for the Americans. This interpretation of the observed spectral differences is consistent with the traditional view of auditory phonetics that French /u/ is produced with a more "peripheral" (i.e., posterior) tongue position (see Flege and Hillenbrand, 1984).

The native English speakers produced /i/ with a significantly higher F2 but a significantly lower F3 than the native speakers of French. Since the auditory system may not resolve formants as close as the average F2 and F3 measured here (Chistovich et al., 1979), the subjects may have perceived an F2–F3 resonance to be higher in frequency for the /i/ produced by the French than American talkers. This is also consistent with the traditional auditory analysis of French /i/ as being more "tense" or "peripheral" (i.e., anterior) than its English cognate.

Another possible cue for accent in the English vowels spoken by the French talkers may have been diphthongization. Vowels like /i/ and /u/ tend to be diphthongized to a greater extent in English than French. It is claimed that Americans' maintenance of diphthongization is one basis for an English accent in French (see, e.g., Politzer and Weiss, 1969; Walz, 1979). The magnitude of dynamic spectral changes in the vowels being considered here was not assessed. However, a careful perceptual appraisal by the author suggested that there was no obvious difference between the native and non-native talkers in degree of diphthongization, probably as the result of the relatively brief duration of their vowels.

It is possible, however, that listeners were responding in part to differences in vowel duration between the native and non-native talkers. The /u/ produced by the native speakers was significantly longer (by an average 25 ms) than that of the French talkers. Their /i/ was a nonsignificant 8 ms longer than the /i/ produced by French native speakers.

There were also temporal differences in the aperiodic portions of syllables. Recall that the /t/ segments extended from the release burst to the onset of periodicity, which is equivalent to VOT as it is traditionally measured. The French talkers produced /t/ with VOT values that were a great deal longer than the VOT measured in French words produced by French monolinguals (Caramazza and Yeni-Komshian, 1974). However, their VOT was shorter than that of the native English speakers. The /t/ edited from French-produced /tu/ syllables averaged about 30 ms shorter than that of the native English speakers, and about 15 ms shorter in /ti/. Both differences were significant (see Table V), which is consistent with other studies of foreign language speech production (Caramazza et al., 1973; Flege, 1980; Flege and Hillenbrand, 1984). There is independent

<table>
<thead>
<tr>
<th>Hybrid syllable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>/u/ + constant 2/</td>
<td>31</td>
<td>30</td>
<td>30</td>
<td>37</td>
<td>37</td>
<td>38</td>
<td>38</td>
<td>33</td>
<td>33</td>
<td>34</td>
<td>38</td>
<td>38</td>
<td>35</td>
<td>35</td>
<td>33</td>
</tr>
<tr>
<td>/u/</td>
<td>28</td>
<td>28</td>
<td>23</td>
<td>23</td>
<td>31</td>
<td>31</td>
<td>26</td>
<td>26</td>
<td>36</td>
<td>36</td>
<td>28</td>
<td>28</td>
<td>31</td>
<td>31</td>
<td>20</td>
</tr>
<tr>
<td>/u/ + constant</td>
<td>2/</td>
<td>31</td>
<td>30</td>
<td>30</td>
<td>37</td>
<td>37</td>
<td>38</td>
<td>38</td>
<td>33</td>
<td>33</td>
<td>34</td>
<td>38</td>
<td>38</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>constant /u/ + /u/</td>
<td>28</td>
<td>28</td>
<td>23</td>
<td>23</td>
<td>31</td>
<td>31</td>
<td>26</td>
<td>26</td>
<td>36</td>
<td>36</td>
<td>28</td>
<td>28</td>
<td>31</td>
<td>31</td>
<td>20</td>
</tr>
</tbody>
</table>

**Table V:** The percentage of times hybrid CV syllables produced by native and non-native speakers were identified as having been produced by a "non-native" speaker (see text for details). Each percentage is based on 100 paired comparison judgments.
TABLE V. Summary of spectral and temporal measurements of the 16 /tu/ and /ti/ syllables produced by native (American) and non-native (French) talkers. Duration is expressed in ms, the center frequencies of $F_1$ – $F_3$ in Hz. Standard deviations are in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>American /tu/</th>
<th>French /tu/</th>
<th>$p$</th>
<th>American /ti/</th>
<th>French /ti/</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$</td>
<td>456(52)</td>
<td>461(123)</td>
<td>0.17</td>
<td>451(38)</td>
<td>431(41)</td>
<td>1.56</td>
</tr>
<tr>
<td>$F_2$</td>
<td>1642(213)</td>
<td>1406(187)</td>
<td>10.9</td>
<td>2694(83)</td>
<td>2490(138)</td>
<td>25.7</td>
</tr>
<tr>
<td>$F_3$</td>
<td>2824(124)</td>
<td>2592(236)</td>
<td>12.1</td>
<td>3165(100)</td>
<td>3421(230)</td>
<td>16.6</td>
</tr>
</tbody>
</table>

Vowel duration 137(22) 112(24) 8.84 0.006 107(31) 100(23) 0.65 0.569
VOT 78(14) 46(17) 32.1 0.000 79(17) 63(19) 6.69 0.014

VI. EXPERIMENT 5

Traditional phonetic accounts describe the /t/ of French as dental and the /t/ of English as alveolar. However, we cannot be certain of how /t/ is actually articulated in either language. Wood's (1975) cineradiographic data indicate, contrary to expectation, a clearly dental place of articulation for the /t/ produced by a single native speaker of (British) English. Data reported by Bladon and Nolan (1977) indicate that (British) English speakers produce /t/ with both a dental and alveolar place of articulation. The same may be true for French (Valdman, 1976).

Another unresolved issue concerns the shape of the tongue during /t/ production. Wenk (1979a,b) indicates that most native French speakers produce /t/ by approximating the tongue tip to the lower incisors and broadly contacting the blade of the tongue against the upper incisors and the alveolar ridge. Wood's (1975) data show blade (i.e., laminar) contact in the /t/ produced by a single English speaker, although most traditional accounts refer to English /t/ as a "tongue-tip" consonant. Bladon and Nolan's (1977) data indicate both "blade" and "tip" articulations of (British) English /t/.

Since we cannot be certain of the articulatory differences between French and English /t/, it is equally uncertain how the /t/ produced by native and French speakers may differ acoustically. However, there are four a priori reasons to think that it might be difficult for an English-speaking listener to detect the presence of a "French" /t/ in English words. First, researchers concerned with foreign language acquisition (e.g., Lado, 1957; Brière, 1966, footnote 4) agree that small differences in the place of contact and/or tongue shape for /t/ are "unimportant" or "hardly detectable." Support for this is provided by Johansson (1978), who reports that native English speakers gave much higher subjective ratings to dental stops produced in English words by Swedish native speakers than to other Swedish-accented consonants.

Second, few languages maintain a contrast between dental and alveolar consonants, implying that such a contrast may be difficult to perceive. (An alternate possibility, of course, is that such a distinction is difficult to consistently articulate.) In languages where more than one phonetic category exists in the dental–alveolar region, a difference in tongue shape (e.g., laminar versus apical; see Chomsky and Halle, 1968, p. 313) may exist alongside a place of articulation difference, or the contrast may be restricted to just a few phonetic environments.

Third, Steven's quantal theory (1973) predicts that variations in the place of articulation in the dental–alveolar region will yield only minimal acoustic differences. This expectation was supported by Lahiri and Blumstein (1981), who report similar spectral properties for the release phase of dental and alveolar stops in Malayalam.

Fourth, as discussed in the Introduction, listeners may "filter out" cross-language phonetic differences such as those which may potentially distinguish the /t/ produced by native and French speakers of English (Trubetzkoy, 1939/69).

There are nonetheless several reasons to think that native English-speaking listeners will be able to distinguish a French from an English /t/. First, Flege and Hammond (1982) found that talkers were able to mimic the VOT differences between native and Spanish speakers of English. Like the potential spectral differences arising from differences in lingual articulation between French and English /t/, this VOT difference is subcategorical in nature. A Spanish-produced /t/ with short-lag VOT values is likely to be heard by English-speaking listeners as /θ/ rather than /d/ (see Lisker, 1978). Second, anecdotal evidence suggests that listeners can also detect purely spectral differences in /t/ production. Jones (1972, p. 142) indicates that the "laminar" (i.e., blade) articulation characteristic of French-accented /t/ produces a "very unnatural effect" in English. Wood (1975) reports that Swedes are "very conscious" of the difference between the (nominally) alveolar /t/ of English and the (nominally) dental /t/ of Swedish, and that this aspect of accented Swedish is considered by Swedes to be a "typical ingredient of an..."
English accent” (p. 174). Third, we saw earlier that listeners were able to detect accent on the basis of subcategorical differences in the /t/ produced by non-native speakers.

Since there were both temporal and spectral differences that might have cued accent in experiment 4, the temporal differences between stops produced by the American and French talkers were neutralized. Only the first 30 ms of each syllable (i.e., just the initial portion of /t/) was presented to subjects in a paired comparison task. Listeners can perceive some differences in place of articulation (e.g., labial versus alveolar or alveolar versus velar) on the basis of the rapid spectral changes found in the initial 10–30 ms of CV syllables (Blumstein and Stevens, 1980). Thus if the French talkers produced /t/ with a dental rather than alveolar place of articulation, and if the distinction between these two kinds of /t/ is auditorily detectable, subjects should be able to detect accent in these very short stimuli.

A. Method

1. Subjects

The listeners were nine female students in speech-language pathology with a mean age of 23.4 years (sd = 2.5 years) who were paid for their participation. One of the nine listeners had participated in experiment 4 about ten days earlier. All were native speakers of American English whose native dialect could be described as “general American.”

2. Speech materials

The stimuli in this experiment consisted of the first 30 ms (± 1 ms) edited from the /tu/ syllables used in experiment 3. These truncated “/t/-burst” stimuli included the burst, frication, and a variable portion of the aspiration-filled interval following stop release. Since the VOT intervals in all the syllables were greater than 30 ms, no periodic energy from the following “vowel” interval was included in any of the 32 stimuli (2 groups X 8 talkers X 2 replicate productions of /tu/). In addition to neutralizing temporal differences between stimuli by fixing the duration of stimuli at 30 ms, rms intensity differences between stimuli were also digitally neutralized, as in the preceding experiments. It should be pointed out that the /t/-burst stimuli used here did not sound very speechlike.

3. Procedure

The /t/-burst stimuli were blocked according to the phrase (“two little dogs” or “two little birds”) in which they had originally been produced. Within a block, the stimuli produced by each of the eight French talkers were paired with each of the stimuli produced by the eight American talkers, twice in the first position and twice in the second position. The interstimulus interval was 1 s; the intertrial interval was set at a minimum of 1 s. This yielded a total of 288 paired comparison judgments for each of the /t/-burst stimuli (8 pairings X 4 presentations X 9 listeners).

The subjects were informed they would hear just the very beginning of the word “two” produced by native and non-native (French) speakers of English. They were told to evaluate each stimulus in terms of whether it resembled the /t/ in “two.” The listeners were instructed to push the left or right button on a response box, depending on whether they judged the first or second member of the stimulus pair to be least likely to have been produced by a fellow native English speaker. No instruction, feedback, familiarization, or training was given before the experiment.

Two analyses were performed. In the “talker” analysis, speaker group (American versus French talkers) was a between-subjects factor, and token (bursts edited from phrase #4 or #5) was a repeated measure. In the “listener” analysis, both speaker group and token were repeated measures.

B. Results and discussion

The results presented in Table VI indicate that spectral information alone suffices to cue accent in the /t/-burst produced by French speakers of English. Overall, 69% of the /t/-bursts produced by native speakers of French were correctly chosen as the “non-native” member of the pair, as against 31% of the American-produced /t/-bursts. The /t/-bursts produced by the American talkers were incorrectly chosen as non-native from 24%-40% of the time. The /t/-bursts produced by the French talkers were chosen as the “non-native” member of the pair 42%-80% of the time. The effect of speaker group was highly significant in both the talker analysis [F(1,14) = 52.2, p < 0.0001] and in the listener analysis [F(1,8) = 17.7, p = 0.003].2 The effect of token did not reach significance.

There were substantial differences among the nine subjects according to the frequency with which /t/-bursts produced by the French talkers were correctly chosen as the “non-native” member of the pair. In decreasing order, the rate of correct accent detection for each listener was: 86%, 86%, 78%, 74%, 65%, 64%, 59%, 58%, and 48%. Chi-square analyses indicated that all but the last subject chose French-produced /t/-bursts as the “non-native” member of

| TABLE VI. The percentage of times the “/t/-burst” stimuli produced by native and non-native speakers of English were chosen as “non-native” in a two-interval forced-choice task. Each percentage is based on 288 paired comparisons. |
|-----------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|                                  | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 2x  |      |      |      |      |      |
| American talkers                 |     |     |     |     |     |     |     |     |     |      |      |      |      |      |      |
| /t/-burst #1                     | 27  | 30  | 23  | 27  | 34  | 38  | 40  | 30  | 31  | 60   | 80  | 74  | 81  | 76  | 74  |
| /t/-burst #2                     | 23  | 34  | 32  | 21  | 34  | 42  | 40  | 28  | 32  | 49   | 69  | 73  | 79  | 77  | 82  |
|                                  | 25  | 32  | 27  | 24  | 34  | 40  | 40  | 29  | 31  | 55   | 75  | 74  | 80  | 77  | 78  |
| French talkers                   |     |     |     |     |     |     |     |     |     |      |      |      |      |      |      |
|                                  |     |     |     |     |     |     |     |     |     |      |      |      |      |      |      |
|                                  | 2x  | 25  | 32  | 27  | 24  | 34  | 40  | 40  | 29  | 31  | 55  | 75  | 74  | 80  | 77  |

James Emil Flege: Detection of French accent 702
been produced with lower frequency than the American-seemed to have a "whistling" quality, but most seem to have lancs that were most often chosen as "non-native" (i.e., those of French talkers 2–7) appear to have been by the author suggested the following. The 12 French-produced /t/-bursts that were most often chosen as "non-native" (i.e., those of French talkers 2–7) appear to have been produced with "laminal" tongue shape (i.e., a relatively long antero-posterior area of the tongue–palate contact). Several seemed to have a "whistling" quality, but most seem to have been produced with lower frequency than the American-produced /t/-bursts.

Many of the accented /t/-burst stimuli contained a perceptual trace of the following vowel (/u/) unlike the /t/-burst stimuli produced by the least "accented" French talkers (#1 and #8) and the American talkers. Lubker and Gay (1982) report that Swedes begin rounding the lips sooner before the acoustic onset of a rounded vowel than Americans, and do so more forcefully. This suggests the possibility that French talkers in this study produced /t/ with greater (or earlier) coarticulated lip rounding than the American talkers. If so, this would be expected to lower the spectrum of the /t/-burst stimuli in comparison /t/-burst stimuli produced with less (or later) lip rounding.

To test this hypothesis, the /t/-burst stimuli were passed through a 32-channel filter bank with pre-emphasis (Voiceprint model 500). Figure 3 presents the average amplitude (db) of energy in the mid-frequency range 970–1555 Hz (filters 8–15), and in the high-frequency range 1555–4555 Hz (filters 16–30) was computed for each of the 32 /t/-burst stimuli. For the three sets of stimuli displayed in Fig. 3, average amplitude (in db) in the two frequency ranges is found in Table VII. Note that for the "accented" French /t/-bursts, average amplitude was about 5 db greater in the mid-frequency range than high-frequency range, whereas for the "unaccented" French-produced /t/-bursts just the reverse was true.

To determine if the relative amplitude of high- compared to mid-frequency energy was related to foreign accent judgments, a Spearman rank-order correlation analysis examining the percentage of times each of the 32 stimuli was chosen as "non-native" and the ratio of high- to mid-frequency amplitude was performed. It revealed that these two variables were significantly correlated ($\rho = 0.43, df = 30, p < 0.001$). The greater the high- to mid-frequency amplitude ratio, the less likely were stimuli to be chosen as "non-native."

From experiment 5 we conclude that differences between native and non-native speakers in the rapid spectral change accompanying release of a prevocalic /t/ are sufficient to cue foreign accent. At present it is unclear whether these spectral differences result from differences in labial coarticulation, or from cross-language differences in the tongue shape/placement associated with /t/ production.

![FIG. 3. Histograms representing the average amplitude, in dB (arbitrary reference), passed by a 32-channel filter bank; (a) demonstrates the pre-emphasis applied; (b) represents 16 /t/-burst stimuli produced by native speakers of American English; (c) represents the 12 French-produced stimuli judged to be accented; and (d) is the average of four French-produced stimuli that evaded detection as accented.](image-url)

**TABLE VII. Average amplitude for the two frequency ranges in Fig. 3.**

<table>
<thead>
<tr>
<th>970–1555 Hz</th>
<th>1555–4555 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native English [Fig. 3(b)]</td>
<td>83.0(2)</td>
</tr>
<tr>
<td>Accented French [Fig. 3(c)]</td>
<td>88.5(4)</td>
</tr>
<tr>
<td>Unaccented French [Fig. 3(d)]</td>
<td>83.1(2)</td>
</tr>
</tbody>
</table>
Given the theoretical importance of determining whether dental and alveolar stops are cued by the same invariant spectral correlates (Lahiri and Blumstein, 1981), it would be useful to perform a similar experiment examining accent detection in French and American-produced stops edited from syllables with an unrounded vowel, such as /ti/.

VII. GENERAL DISCUSSION

This series of experiments demonstrates that human listeners are acutely sensitive to divergences from the phonetic norms of their native language, such as those which lead to the perception of foreign accent. The average rate at which accent was detected in short stretches of English spoken by non-native (French) talkers varied between 63% and 95% in forced-choice and paired-comparison tasks. Accent detection appeared to be somewhat better for relatively long compared to short excerpts of speech, yet remained reliable as the speech samples were progressively reduced from phrase to syllable to segment size. In the final experiment, listeners accurately detected accent when presented with just a fixed portion of a phonetic segment (the first 30 ms of the syllable /tu/). However, because stimuli of different lengths were presented to different groups of subjects, we can draw no firm conclusion concerning whether accent is more readily detectable in syllables compared to individual phonetic segments or parts of segments.

One conclusion to be drawn from this study is that foreign accent detection does not depend on experience with a specific variety of accented speech. Experiment 3 showed that listeners who were unfamiliar with French-accented English accurately detected accent in the English spoken by native speakers of French in a forced-choice task. This does not imply, of course, that the listeners recognized the language background of the non-native speakers. This may require a great deal of experience. It remains to be determined by future research whether specific varieties of accented speech (e.g., Italian- or German-accented English) can be recognized. It is likely that correct accent recognition (as opposed to detection) would require longer stretches of speech than those examined here.

Another finding was that accent was detected equally well in speech produced by the non-native speakers in a phrase reading task and in a spontaneous speech task. This suggests that "attention to speech" (see Flege and Hillebrand, 1984) does not affect the authenticity with which non-native speakers produce phonetic segments or syllables in a foreign language.

Experiments 2–5 provided ample evidence that subcategorical phonetic differences between native and non-native speakers are detectable, at least under ideal listening conditions in which listeners can focus attention on very short stretches of speech. Few of the verifiably accented syllables produced by French native speakers in this study differed from English syllables in an obvious way. Their "non-native" characteristics could not be easily described using IPA symbols or diacritics, at least by the author. Each would be transcribed as [tu] or [ti] using broad transcriplional categories. Listeners were nonetheless able to detect accent in these syllables, apparently as the result of minimal spectral differences in /u/ and /i/, and as the result of differences in lingual articulation or labial coarticulation associated with the production of prevocalic /i/.

The present results are important in demonstrating that, in the perception of naturally produced speech, the listener does not irreversibly "filter out" subcategorical differences which are insufficient to change phonetic category identity. At least some of the phonetic differences that may distinguish phonetically different cognate sounds in two languages (e.g., the /u/ found in French and English) are accessible to the untrained listener. This means that the native French speaker who maintains French patterns of segmental articulation in English words does not necessarily do so because French–English phonetic differences are auditorily inaccessible (cf. Flege and Hammond, 1982). Cross-language phonetic interference (Flege and Port, 1981) may instead stem from difficulty in modifying pre-established patterns of articulation, or from the fact that information deriving from the detection of cross-language phonetic differences does not readily influence central phonetic representations (Flege, 1981).

It is possible that even if listeners do not consciously detect subcategorical phonetic differences between native and non-native speakers in ordinary circumstances, such differences nonetheless contribute to the detection of accent. Morse et al. (1977) found that listeners who did not consciously hear a category shift from /ba/ to /ga/ in a synthetic speech sound continuum nonetheless manifested physiological evidence (i.e., heart rate changes) that such a shift had been detected. Whalen (1982), using techniques similar to those of experiment 4, found that subjects were significantly slower in identifying consonants appended to a vowel they had not originally been coproduced with than phonemes occurring in their original context, despite the fact that the category identity of the consonants was unaltered by the process of cross splicing. These results suggest that listeners evaluate everything they hear in speech stimuli, even acoustic dimensions that are not overtly detectable and/or do not affect category identity.

This inference is supported by an incidental finding of this study. One French talker's production of the phrase "two little dogs" was incorrectly accepted as "native," even though the first syllable (/tu/) was identified as "non-native" when presented in isolation in a later experiment. Thus auditorily detectable differences in a phonetic segment deriving from cross-language phonetic interference may not always be consciously perceived when attention is spread over many phonetic segments and suprasegmental dimensions in a relatively long stretch of speech.

The study leaves unanswered many basic questions concerning how accent is detected. One hypothesis is that accent is detected whenever the listener detects a pronunciation difference that is sufficient to alter phonetic category identity. Controlled laboratory studies (e.g., Cole, 1981) demonstrate that listeners can detect sound "substitutions" that have been intentionally introduced into connected speech. It should be pointed out, however, that subjects in such experiments are expressly listening for mispronunciations rather than for meaning, as in normal speech communi-
cation. The degree of perceived accent might follow directly from the number of detectable sound substitutions (Ryan et al., 1977).

A phonetic transcription by the author of the phrases examined in experiment I revealed the "substitution" or "deletion" of a number of phonetic segments by the non-native speakers. However, it is unclear whether all such mispronunciations contribute equally to the detection, recognition, or evaluation (Brennan and Brennan, 1981a, b) of accented speech. One unresolved question concerns the effect of prior experience and expectation. Labov (1972) observes that some New York City residents overtly stigmatize pronunciations such as "deese" for "these." This implies that listeners may be especially sensitive to pronunciation differences they have come to expect. However, it might not be typical of foreign accent perception since it represents tacit knowledge of patterned variation in pronunciation that carries social meaning.

There are other reasons to think that listeners might be less sensitive to differences they have come to expect. The ability to learn patterns of correspondence between varieties of speech (Lovins, 1976) seems to be the basis for why unfamiliar children appear to become rapidly more intelligible to adult native speakers after some amount of exposure. The possibility exists that listeners learn to ignore predictable changes in pronunciation (see Gibson, 1969), whereas they detect accent in relatively less predictable pronunciation differences, such as the Cuban's pronunciation of "vase" as "base." A diminished sensitivity to the relatively more "predictable" pronunciation differences between adults and children (such as "wabbit" for "rabbit") may explain why children are not usually considered to speak with an accent. Differences between native and non-native speakers, on the other hand, may lead to the perception of accent because they are less predictable, owing to the wide range of phonetic and phonological differences between languages.

Another unanswered question is whether listeners are more sensitive to categorical than to subcategorical phonetic differences. For example, will the perceived "substitution" of /I/ for /I/ in a French pronunciation of "little" be detected more easily, or weigh more heavily, than the production of English /I/ with short-lag VOT values (see Johansson, 1978)? One hypothesis that warrants investigation is that categorical differences in pronunciation between native and non-native speakers which lead to the perception of unintended—but possible—lexical items will be less detectable than those leading to the perception of nonwords. Another hypothesis is that subcategorical differences are more easily detected in familiar or expected words than in relatively less familiar or unexpected words. This is in keeping with the general facilitation of "feature" detection by the presence of context which characterizes subjects' conscious reports of sensory experiences (see, e.g., Rubin et al., 1976).

The detection of accent probably depends on phonetic rather than auditory perception, for auditorily detectable differences from the norm such as hoarseness or nasality are not interpreted as accent. This inference is consistent with Nooteboom's (1973) hypothesis that the phonetic representations stored in long-term memory contain a great deal more language-specific detail concerning segmental articulation than is needed to simply identify phonetic categories.

The construct of a phonetic category "prototype" might profitably be applied to the problem of foreign accent detection. As developed by Rosch (e.g., 1973, 1978), a prototype model posits that objects (e.g., colors, birds, pieces of furniture) are categorized on the basis of comparison to internal prototypes which represent a category's core properties, rather than in terms of decision processes involving the boundaries between categories. Prototypes are often developed through experience with many members of a category (although physiological mechanisms underpinning sensation are undoubtedly important for categorization, especially in initial stages of development; Kuhl, 1980). Specific exemplars of a category may never possess all and only the properties of the category prototype. Instead, they are accepted or rejected as belonging to the category on the basis of how closely they conform to the prototype.

In recent years the prototype construct has been extended to speech research. Oden and Massaro (1978) and Massaro and Oden (1980) proposed that phonemes are identified by comparing speech stimuli to phonetic category prototypes stored in long-term memory (cf. Repp, 1976; Miller, 1977; Samuel, 1982). Their model posits that prototypes reflect the influence of linguistic experience, representing a configuration of ideal relative values of many independently perceivable acoustic dimensions for sounds found in specific syllable environments. Values for each auditory-acoustic parameter are integrated according to a simple metric, and the relative proximity of a stimulus to possible prototype "candidates" is evaluated. The stimulus is then identified in terms of the closest prototype.

If phonetic categories are represented by prototypes, we have some basis for understanding experience-based increases in speech recognition and foreign accent detection. One hypothesis for why experienced listeners seem to recognize degraded speech signals better than less experienced listeners, and detect accent more readily, is that they have developed more highly elaborated category prototypes (which include both variant and invariant stimulus properties) against which to judge stimuli. The ability to perceive significant departures from prototypes—including accent—may increase as category prototypes are better defined through exposure to the wide range of possible phonetic realizations of phonetic categories presented in the speech of many talkers. This prediction must be rigorously tested in future research. However, the existence of detailed phonetic "prototypes" seems to be well supported by the present demonstration of human sensitivity to accent.

Acknowledgments

This study was supported in part by an NIH grant (NS07107) to the Department of Communicative Disorders at Northwestern University, and by a grant-in-aid provided by the Graduate School of Northwestern University. The author would like to thank members of the French departments of Northwestern University and Indiana University who participated as talkers; J. Hillenbrand and K. Stevens for help with the acoustic analyses; C. Prall and S. Smith for...
software support; and S. Fletcher, T. Gottfried, and P. Lade-
folded for their comments on a previous version of this ar-

1Empirical support for "phonological filtering" is provided by a cross-lang-
guage study of vowel discrimination. Jacob et al. (1980) found that native
English speakers had no trouble discriminating naturally produced pairs of
vowels (e.g., /i/-/a/, /u/-/u/) embedded in a CVC context. Hebrew-
speaking listeners also easily discriminated stimulus pairs in which both
vowels (e.g., /i/-/u/), or neither vowel (e.g., /a/-/u/), occurred in their native
language. However, the Hebrew subjects made many errors in discriminating
pairs in which just one vowel had phonetic status in Hebrew (e.g., /i/-/
/u/ or /u/-/u/). This was probably due to their categorization of an unin-
liair vowel (e.g., /u/-/u/) as a familiar vowel found in Hebrew (e.g., /i/-/i/)
Like Trubetzkoy, the authors concluded that linguistic experience
"forces" the listener to process "only those acoustic cues which carry some
degree of linguistic relevance" in the native language.

One listener reversed the labeling of native and non-native talkers, choos-
ing the American-produced /t/-bursts as the "non-native" member of the pair
74% (378/511) of the time. This reversal has been corrected in Table
VI since no training procedure aimed at preventing such reversals was un-
taken. If this subject's reverse labeling had not been corrected, the over-
all rate of correct detections would have decreased from about 69% to
63%. It should be pointed out that although correct detections would still
exceed false alarms by a ratio of about 2:1, the effect of speaker group in the
"listener" analysis would no longer reach significance [F(1,8) = 4.33, p
< 0.069].

This possibility could be tested by playing the /t/-burst stimuli to monolin-
gual speakers of English and French, and to bilingual French-English
speakers. If a listener's judgment of what constitutes a "good" /t/ is modi-
fied by exposure to two systematically different pronunciations of /t/ we
would expect the following: monolingual native French-speaking listeners
should judge the /t/-burst stimuli produced by American talkers to be "ac-
cented." Monolingual native English-speaking listeners should judge the
stop produced by the French talkers to be "accented." The bilingual listen-
ers, on the other hand, should give the appearance of being unable to detect
accent. That is, they should be unable to reliably distinguish between the
two groups of talkers.

The center frequencies of channels 1-8 (200-1080 Hz) were separated by
110 Hz, channels 9-15 (1050-1955 Hz) by 125 Hz, channels 16-22 (1955-
3005 Hz) by 150 Hz, channels 23-28 (3005-4555 Hz) by 175 Hz, and chan-
els 29-38 (4555-7500 Hz) by 200 Hz.

For example, word-final /z/ (in "dogs" and "birds") was frequently de-
voiced (i.e., heard as /s/) by American and French talkers (10/24 and 13/
24 of observed instances, respectively). The native French talkers produced
/z/ so that it was perceived as /s/ in 16/24 productions of "little," whereas
the native English speakers never did so. Five of the six French talkers
produced "bird" with a French-like vocalic /r/ rather than the dorsal ap-
tip and blade alveolars in English. Unlike the text-dependent phonetic var-
ation of /r/ by the remaining French talker, this does not represent a con-
text-dependent phonetic variation present in American English dialects.
Also observed was the apparent deletion of /d/ in /birds/ (by two French
talkers) and the /s/ from "dogs" (by one French talker). Another "segmental-
" difference between native and non-native speakers concerned produc-
tion of post-stressed /u/. The native French talkers produced the /u/ in "up"
[([13/24)] as a full stop rather than as a flap in 11 of 24 instances, whereas
the American talkers flapped /u/ in every instance (24/24).

Sved (1981) presented 8-s excerpts of speech produced by ten native and
ten non-native English speakers to several groups of listeners. He noted
higher rates of accent identification for older compared to younger English
speaking children, for native compared to non-native English-speakers,
and for advanced compared to beginning speakers of English as a second
language. McCarr [1981] found that listeners who are familiar with "deaf-
accented English" comprehend more of it than listeners unfamiliar with
the speech of the deaf. Elliott (1979) found that older children were better
able than younger children to recognize words in a predictable context
when the speech signal was degraded by multtalker babbling. Parnell and
Amerman (1978) found that older children and adults were better able than
younger children to recognize CV syllables that had been altered by elec-
tronic gating techniques.

continuum: Cross-language tests," Proceedings of the 6th International
Asher, J., and Garcia, R. (1999). "The optimal age to learn a language,
Beebe, L. (1980). "Sociolinguistic variation and style shifting in second
Bloomfield, L. (1933). Language, Holt, Rinehart, and Winston, New York,
spectra for stop consonants in different vowel environments," J. Acoust.
Soc. Am. 67, 648-662.
Brennan, E., and Brennan, J. (1981a). "Accent scaling and language at-
24, 207-221.
Brennan, E., and Brennan, J. (1981b). "Measurements of accent and at-
titude toward Mexican American speech," J. Psycholinguist. Res. 10, 487-
501.
tedness by magnitude estimation and sensory modality matching," J.
42, 769-796.
"The acquisition of a new phonological contrast: The case of stop conson-
Chistovich, L., Shelkin, R., and Lubinskiaga, V. (1979). "Centres of gra-
vity and spectral peaks as the determinants of vowel quality," in Front-
tiers of Speech Communication Research, edited by B. Lindblom and S.
and Row, New York).
Native Language and Foreign Language Acquisition, edited by H. Winitz
Dickerson, L. (1975). "The learner's inter-language as a system of variable
Elliott, L. (1979). "Performance of children aged 9 to 17 years on a test of
speech intelligibility in noise using sentence material with controlled
Feige, J. (1980). "Phonetic approximation in second language acquisition,
Feige, J. (1981). "The phonological basis of foreign accent: A hypothesis,
"TOD, 15, 443-449.
Human Communication and Its Disorders, Vol. 1, edited by H. Winitz
differences between language varieties," Stud. Second Lang. Acquis. 5,
1-18.
Feige, J., and Hillenbrand J. (1984). "Perceptual limits on phonetic accura-
cy in foreign language speech production," J. Acoust. Soc. Am. 76 (to
appear).
bic to English," Lang. Speech 24, 125-146.
Univ. of Gotenburg, Sweden).
bridge, U.K.).
ies for Anthony Cohen, edited by M. van den Broecke, V. van Heuven, and
infancy," in Child Phonology, Vol. 3, Perception, edited by G. Yeni-Kom-
Labov, W. (1972). Sociolinguistic Patterns (Univ. of Pennsylvania, Phila-
}


Wenck, B. (1979b). *Personal communication.*

