

Limits on phonetic accuracy in foreign language speech production

James Emil Flege

Department of Biocommunication, University of Alabama in Birmingham, University Station, Birmingham, Alabama 35209

James Hillenbrand

Department of Communicative Disorders, Northwestern University, 2209 Sheridan Road, Evanston, Illinois 32601

(Received 20 June 1983; accepted for publication 27 April 1984)

This study examined the French syllables /tu/ ("tous") and /ty/ ("tu") produced in three speaking tasks by native speakers of American English and French talkers living in the U. S. In a paired-comparison task listeners correctly identified more of the vowels produced by French than American talkers, and more vowels produced by experienced than inexperienced American speakers of French. An acoustic analysis revealed that the American talkers produced /u/ with significantly higher F_2 values than the French talkers, but produced /y/ with F_2 values equal to those of the French talkers. A labeling task revealed that the /y/ vowels produced by the experienced and inexperienced Americans were identified equally well, but that the experienced Americans produced a more identifiable /u/ than the inexperienced Americans. It is hypothesized that English speakers learn French /y/ rapidly because this vowel is not—like French /u/—judged to be equivalent to a vowel of English. The French and American talkers produced /t/ with equal VOT values of about 55 ms, which is intermediate to values commonly observed for monolingual speakers of French and English. It is hypothesized that the bilingual talkers judged the /t/ of French and English to be equivalent, which affected their perceptual target for French /t/ and ultimately their production of this stop.

PACS numbers: 43.70.Ve, 43.70.Dn, 43.70.Gr, 43.70.Bk

INTRODUCTION

Most adults who learn a foreign language speak with an "accent" which derives in part from phonological and phonetic differences between their native language (L1) and the target foreign language (L2). This raises the question of the extent to which humans are capable of learning new speech patterns, or modifying existing ones, once L1 has been thoroughly acquired. This study examined factors that might limit the accuracy with which adult learners produce the phones (speech sounds) found in a foreign language.

Researchers (e.g., Lado, 1957) have frequently observed that L2 learners tend to produce the words of a foreign language with phones that can be heard in the phonetic surface of L1. Weinreich (1953) hypothesized that this resulted from the "interlingual identification" of L1 and L2 phones. Valdman (1976, p. 38) observed that:

...the articulatory habits of the foreign language partially overlap those of the native language... The student must learn to make new responses to stimuli which are interpreted as identical to native language stimuli. For instance, French and English /s/ differ with regard to place of articulation. The former is a dental, and the latter is an alveolar. The partial similarities he perceives in the acoustic signal of French /s/ will lead an English speaker to respond with the alveolar rather than the dental sound.

This represents the hypothesis that when an L2 phone is "identified" with an L1 phone, the L1 phone will be used in place of it (i.e., be "substituted" for it). Such interlingual identification appears to depend on the auditory, and per-

haps articulatory, similarity of L1 and L2 phones. However, empirical research has yet to: (1) objectively quantify the degree of phonetic similarity between phones found in two languages, (2) demonstrate which specific phones in L1 and L2 are "identified" with one another, or (3) determine whether the interlingual identification of L1 and L2 phones will cease as the learner becomes familiar with the phonetic differences that may distinguish L1 and L2 phones.

Some L2 phones have a direct counterpart in L1 with which they can be identified. To a first approximation, this includes L1 and L2 phones that are transcribed using the same IPA symbol. For example, instances of /t/ occurring in French and English words are likely to be regarded by the L2 learner as being different realizations of the same category because of their overall phonetic similarity. The interlingual identification of such pairs might be expected to occur even when acoustic differences that may distinguish the L1 and L2 phones are auditorily detectable (see Flege, 1984a). Judging acoustically different phones to be members of the same category is a fundamental aspect of human speech perception.

Other L2 phones bear less obvious resemblance to phones in L1 and may therefore not be regarded as the realization of any L1 category. For native English speakers, this may include phones such as the front-rounded vowel /y/ of French, the clicks of Southern Bantu languages, and the pharyngeal fricatives of Semitic languages. We will refer to such phones as "new," although it should be apparent that any phone encountered in a foreign language—no matter how exotic—is likely to bear some degree of articulatory and acoustic similarity to phones found in the learner's L1.

One aim of this study was to determine whether L2 learners produce "new" L2 phones more accurately than L2 phones which have a counterpart in L1. Valdman (1976) hypothesized that new L2 phones are learned more easily than L2 phones which have an L1 counterpart because they evade interlingual identification and thus tend not to be produced with previously established patterns of segmental articulation (cf. Koutsoudas and Koutsoudas, 1962). Koo (1972) suggested that talkers do not actually need to "learn" new L2 phones since they can be produced by recombining the features of L1 phones. This implies that new L2 phones and those which have a direct counterpart in L1 will be produced with equal accuracy.

Several previous studies have reported data bearing on this issue (Briere, 1966; Politzer and Weiss, 1969; Johansson, 1973; Walz, 1979; Flege and Port, 1981), but none has specifically tested the hypothesis that new L2 phones are produced more accurately than those with a direct counterpart in L1.¹ Taken as a whole, they suggest that L2 learners produce new L2 phones *less* well than L2 phones with an L1 counterpart. However the evidence which now exists is insufficient to confirm or disconfirm the hypothesis. Most of the previous studies examined the speech production of talkers with little or no knowledge of the foreign language from which the L2 phones were taken; most examined mimicry rather than spontaneous speech production; and most relied solely on phonetic transcription, often by non-native speakers.

We applied perceptual and acoustic criteria to the French syllables /tu/ ("tous") and /ty/ ("tu") to objectively examine adults' production of L2 phones.² English /u/ appears to be produced with significantly lower *F*₂ values than its French counterpart.³ French /y/, on the other hand, is a "new" phone that has no direct counterpart in English. Contrastive analysis (e.g., Le Bras, 1981) predicts that L2 learners will replace a new L2 phone with a "close" L1 phone, but offers no satisfactory method for determining which L1 phone is *closest* to the new L2 phone.⁴ If a new L2 phone is consistently replaced by a single L1 phone the possibility exists that the L2 phone has been identified with that L1 phone. American speakers of French often realize French /y/ as an /u/-quality vowel (Gaudin, 1953; Walz, 1979). If both the /y/ and /u/ of French is identified with English /u/, native English speakers might appear to produce French /u/ more accurately than French /y/ since, in that case, they would be likely to "substitute" English /u/ for both vowels. Another possibility is that native English speakers do not identify French /y/ with any vowel category of English and therefore learn to produce French /y/ without reference to speech patterns established for the articulation of English. If so, French /y/ may be produced more accurately than French /u/, at least if adults remain capable of learning to produce new phones.

It seems likely that the amount of French language experience will affect the production of French /u/ and /y/ by native English speakers. Previous studies have focused on speech timing in L2 production. Relatively experienced L2 learners have been observed to produce L2 phones with more nativelike temporal properties than relatively inexper-

enced L2 learners (see Flege, 1984b). A nonauthentic pronunciation of vowels leads to foreign accent (Elsendoorn, 1983) but no study, to our knowledge, has specifically examined the effect of experience on learners' production of L2 vowels. Thus another aim of this study was to determine whether experienced American speakers of French produce French vowels more accurately than less experienced Americans and, if so, whether experience equally effects production of /y/ and /u/.

Most previous L2 research has emphasized the difficulty of establishing new motor plans for L2 phones, or the difficulty in modifying pre-existing ones. The final aim of this study was to test the hypothesis that an important cause of foreign accent is the development by L2 learners of inaccurate *perceptual* targets for L2 phones.⁵ Flege (1981, 1984b) hypothesized that interlingual identification leads the L2 learner to merge the phonetic properties of L1 and L2 phones that have been identified as belonging to the same category. According to this hypothesis, learners' perceptual target for L2 phones may evolve as a function of experience with L2 (see Caramazza *et al.*, 1973; Williams, 1980), but their perceptual target for L2 phones may never match those of L2 native speakers because of interlingual identification. Flege's (1984b) model leads to the prediction that, with sufficient experience, L2 learners will produce stop consonants differently in L2 than L1 (if the L1 and L2 stops differ phonetically), but will never perfectly match native speakers of L2. For example, if English learners judge that the /t/ of English and French belong to the same category, it is predicted they will persist in producing French /t/ with relatively long (English-like) VOT values because their perceptual target for French /t/ will merge the properties of French /t/ (including its short-lag VOT values) and English /t/ (with its long-lag VOT values).

Existing studies support the general prediction that L2 learners will be only partially successful in producing L2 phones. Studies have shown that learners whose L1 realizes /p,t,k/ with short-lag VOT values produce English /p,t,k/ with VOT values that are longer than those characteristic of L1, but are nevertheless too short by English phonetic standards (Port and Mitleb, 1980; Flege and Port, 1981; Caramazza *et al.*, 1973; Williams, 1980; cf. Flege and Hammond, 1982). However, to our knowledge no study has examined the production of short-lag stops in a target foreign language by learners whose L1 realizes /p,t,k/ with *long-lag* VOT values. Kewley-Port and Preston (1974) hypothesized that short-lag stops are less difficult to produce than long-lag stops. It is therefore possible that native English speakers may succeed better in producing the short-lag stops of French (Caramazza and Yeni-Komshian, 1974) than native French speakers produce the *long-lag* stops of English. If Americans accurately produce French /t/ with the short-lag VOT values typical of French monolinguals, it would disconfirm the hypothesis that interlingual identification creates an absolute upper limit on the extent to which L2 learners approximate the phonetic norms of a target foreign language.

Flege's (1984b) model should apply to vowel production as well as to the VOT dimension in stop consonants.

Native English speakers are likely to identify French /u/ with English /u/, despite the fact that the spectral acoustic differences distinguishing these vowels are auditorily detectable (Flege, 1984a). As a result, English learners of French may develop a perceptual target for French /u/ differing from that of French monolinguals because they merge the phonetic properties of French and English /u/. If so, they will at best produce French /u/ with formant values intermediate to the values observed for French and English monolinguals. Should the production of French /u/ by English speakers of French be observed to match that of French monolinguals, it would seriously undermine the importance of interlingual identification as a factor limiting adults' success in producing L2 phones.

One other aspect of the present data will serve to test the importance of interlingual identification. Previous studies (e.g., Flege and Port, 1981) show that the phonetic characteristics of L1 phones are often maintained in L2 speech production. To the best of our knowledge, no previous study has tested the effect of L2 learning on L1 speech production. If identifying an L2 phone with a phone in L1 affects the perceptual target developed for the L2 phone, it should also affect the perceptual target previously established for the L1 phone.⁶ Flege (1981, 1984b) hypothesized that when learners identify an L2 phone with a phone in L1 they will eventually begin producing the L1 phone so that it resembles the counterpart phone in L2. For example, a native French speaker who identifies English /t/ with the /t/ of French should eventually begin producing French /t/ with VOT values that exceed the short-lag VOT values typical of French monolinguals. A failure to note an L2 effect on L1 speech production would also tend to undermine the importance of interlingual identification.

Our study is divided into three parts. In experiment 1 listeners label the French syllables /tu/ and /ty/ produced by native French and English talkers. If new L2 phones are produced more accurately than L2 phones which have a direct counterpart in L1, the /ty/ produced by native English speakers should be correctly labeled more often than /tu/. If experience leads to increased L2 production accuracy, the syllables produced by relatively experienced English speakers of French should be correctly labeled more often than syllables produced by less experienced native English speakers. The effect of experience is further explored in experiment 2, which examines the identifiability of /ty/ and /tu/ in a paired-comparison task.

Finally, we report acoustic measures of VOT and formants 1-3 in the /tu/ and /ty/ syllables that were perceptually tested in experiments 1 and 2. The predicted effect of interlingual identification is that French-English bilinguals and English-French bilinguals will produce the /t/ occurring in French words with VOT values that are intermediate to those observed for monolingual speakers of English and French. Further, native English speakers of French are predicted to produce French /u/ and with *F*₂ values intermediate to those observed for French and English monolinguals.

I. EXPERIMENT 1

This experiment examined the accuracy with which native and non-native talkers produced the French syllables

/tu/ and /ty/. Accuracy was assessed by computing the frequency with which French-speaking listeners correctly identified (i.e., as intended) these syllables. We examined production of /tu/ and /ty/ in three different speaking tasks to ensure a representative sampling of French speech production. To examine the effect of linguistic experience, we compared two groups of native English speakers who differed in overall French language experience.

A. Methods

1. Talkers

Three groups of talkers differing in language background and experience were recruited from a university community for the present study. Each group consisted of seven adult talkers with self-reported normal hearing. The native French-speaking group, designated group NF, consisted of six French women and one Belgian woman with a mean age of 38 years. These talkers had lived an average of 12.2 years in an English-speaking environment (principally Chicago), and four were married to native English speakers.

There were two groups of native English speakers who spoke French as a second language, all women from the Midwest. The talkers in one group, designated group NE-1, consisted of undergraduate students with a mean age of 22 years who had spent the previous academic year in Paris. A more experienced group of native English speakers, designated group NE-2, consisted of women with a mean age of 32 years who held advanced degrees in French and were teaching French at an American university. Talkers in this group had all spent several periods of time in France, the total averaging 1.3 years. One was married to a native French speaker. Talkers in the two American groups, like the native French speakers, were using English as their primary language at the time of the study.

Talkers in both native English-speaking groups began to study French in junior or senior high school between the ages of 11 and 17 years. However, none of them appears to have acquired French as a functional second language prior to about age 20. A language background questionnaire indicated that, compared to talkers in group NE-1, those in group NE-2 had substantially more formal instruction in French language and literature, rated their own production and comprehension of French somewhat higher, and used French somewhat more often on a daily basis in the period immediately preceding the experiment.

No attempt was made to objectively assess the French-speaking proficiency of the two American groups, for the intent was simply to constitute two extreme groups differing in overall experience. There was a clear difference between the two groups in terms of the length of time they had used French to communicate. For talkers in NE-1 this was effectively less than a year, since none of them had used French on a regular basis since their return from Paris 6 months previously. The talkers in group NE-2, on the other hand, had used French on a fairly regular basis for an average of about 10 years. Thus it seems reasonable to refer to the talkers in group NE-1 as "inexperienced," and those in group NE-2 as the relatively "experienced" speakers of French.

2. Speech materials

The following two sets of phrases were used in counter-balanced order to elicit production of the French syllables /tu/ and /ty/ by the native speakers of French and English:

| | |
|--------------------|------------------|
| Tous les prêtres | Tu les montres |
| Tous les évêques | Tu les opposes |
| Tous les soldats | Tu les observes |
| Tous les marins | Tu les renvoies |
| Tous les médecins | Tu les obtiens |
| Tous les dentistes | Tu les informes |
| Tous les gendarmes | Tu les regardes. |

Note that the utterance-initial syllable (/tu/ or /ty/) in both phrase sets was held constant. The sound following the initial syllable (/l/) was held constant across the two phrase sets to make possible a direct comparison between /tu/ and /ty/. The final word in each phrase tended to receive contrastive stress because it varied across phrases.

The two sets of phrases were produced in three progressively more demanding speaking tasks. The first task was simply to read the test phrases found in one set, each preceded by the phrase number and a pause. The next task was to generate an original sentence, initiating each with one of the phrases that had just been produced in isolation. Production was cued by the same written list of seven phrases used in the phrase production task.

The final task was to produce a story based on the seven phrases. The principal requirement was that the story include a complete sentence initiated by each of the seven phrases in the set. The talkers were given a set of cards, each bearing one of the phrases to be used. They were to arrange the cards on the table before them in order to outline their intended story. Talkers were permitted as much time as necessary to silently rehearse. They were permitted to intersperse additional sentences not initiated by one of the test phrases in the story, as needed. The talkers were not required to say the number of the phrase before initiating each utterance, as in the previous two speaking tasks. However, they were told to pause before each sentence to ensure that the syllable of interest (/tu/ or /ty/) occurred in absolute utterance-initial position in the story task as it had in the phrase and sentence tasks.

The three speaking tasks were modeled using a set of seven English phrases. The talkers then practiced the speaking tasks using these English phrases. All but one talker, who was replaced, was able to perform the three speaking tasks satisfactorily. The story produced in the third task was highly natural and spontaneous in the authors' estimation, despite the fact that the talkers were required to pause before each sentence.

The speech material was recorded (Sony model TCD5M) in a sound-treated room with the experimenter seated about 5 ft from the talker. An electret condenser microphone (Nakamichi model CM-300) was positioned about 6 in. from the talker's mouth. To counteract the tendency for talkers to hyperarticulate in the presence of a microphone (Labov, 1972), talkers were told that the experiment examined "language creativity." Debriefing afterwards revealed that none of the talkers were aware the experiment actually

focused on pronunciation, and none attached special significance to the fact that a single syllable (/tu/ or /ty/) recurred at the beginning of each phrase.

3. Stimuli

A total of 252 syllables were edited from the speech material for perceptual analysis (3 groups \times 7 talkers \times 2 syllable types \times 2 replicate productions \times 3 speaking tasks). Phrases #4 and #5 from the two phrase sets were digitized at a 10-kHz sampling rate with 12-bit amplitude resolution. The /tu/ and /ty/ syllables initiating these phrases were then isolated using the segmentation criteria illustrated in Fig. 1. The left cursor was placed about 3 ms to the left of the sharp increase in waveform energy signaling the release of constriction of /t/. The right cursor was placed at the zero crossing that was judged to best demarcate the end of the "vowel" (/u/ or /y/) and the beginning of the following "consonant" (/l/).

Segmentation was based on changes in waveform shape and intensity, together with a perceptual appraisal. Segmentation was based on perceptual appraisal alone in the 5%–10% of syllables in which no discontinuity was visually evident in the waveform. Successive glottal periods were eliminated one at a time from the right of the syllable until "/l/ coloring" was no longer perceptible. The average duration of vowels edited in this way was 48.8 ms (s.d. = 16) for group NF, 44.8 ms (s.d. = 15) for group NE-1, and 51.1 ms (s.d. = 20) for group NE-2.

4. Subjects

The subjects were seven female native speakers of French with a mean age of 38 years. Six had served as talkers about 3 months before the experiment. Debriefing after the experiment indicated that none of these subjects were aware that some of their own syllables had been included among the stimuli presented.

5. Procedures

The /tu/ and /ty/ syllables were stored on a high-speed mass storage device for later on-line presentation to listeners

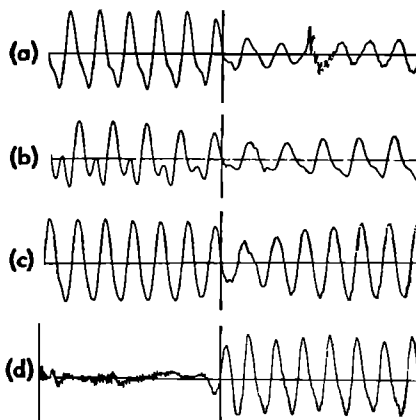


FIG. 1. Panels (a)–(c) illustrate the point in the acoustic waveform of three different /tu/ syllables where /u/ was segmented from the following /l/ sound on the basis of changes in waveform shape and amplitude. The criterion used for the measurement of VOT is illustrated for a /tu/ syllable in (d).

in a two-alternative forced-choice test. The syllables were blocked on speaking task (phrases, sentences, story) according to the phrase (#4 or #5) in which they had been produced. Within a block, each syllable was presented five times. This yielded a total of six blocks, each containing 210 stimuli (21 talkers \times 2 syllables types \times 5 repetitions). The stimuli within each block were digitally normalized for overall rms intensity.

The task was to label each stimulus as "tu" (/ty/) or "tous" (/tu/). The subjects were informed that the syllables had been edited from longer stretches of speech, and that an unspecified number of the talkers were not native speakers of French.

The percentage of times each subject correctly identified the /ty/ and /tu/ syllables was computed separately for each of the three speaker groups (NF, NE-1, NE-2) in each of the three speaking tasks (phrase, sentence, story). The maximum number of correct identifications of /ty/ and /tu/ was 70 (7 talkers \times 2 replicate productions \times 5 presentations). In addition, the percentage of times /tu/ and /ty/ syllables produced by each of the 21 talkers were correctly identified was computed separately for each speaking task. The maximum number of correct identifications was again 70 (7 listeners \times 2 replicate productions \times 5 presentations).

B. Results

Syllables produced by the native speakers of French (NF) were correctly identified more often (about 95% of the time) than syllables produced by either the experienced Americans (about 75% correct) or the inexperienced Americans (about 60%). The /tu/ syllables produced by the native French and experienced Americans were correctly identified at a somewhat higher rate than /ty/. For the inexperienced American talkers, on the other hand, /tu/ was correctly identified at a substantially lower rate (about 45% correct) than /ty/ (about 70% correct). As a result, there was little difference between the experienced and inexperienced American talkers for /ty/, but a substantial difference between these two groups for /tu/.

The percent correct identification scores were transformed using an arcsine transformation (Kirk, 1968, p. 66) because homogeneity of variance cannot be assumed when an analysis is based on percentages. The transformed scores were submitted to a three-way analysis of variance in which speaker group (NF, NE-1, NE-2), speaking task (phrases, sentences, story), and syllable type (/tu/ or /ty/) were all repeated measures.

The effect of speaking task was not significant ($p < 0.01$). However, the interaction of speaker group \times syllable type was highly significant [$F(2,12) = 51.27, p < 0.001$]. Tests of simple main effects indicated that the /tu/ syllables produced by talkers in the native French (NF) and experienced American (NE-2) groups were correctly identified more often than /ty/ syllables, whereas the /tu/ syllables produced by the inexperienced Americans (group NE-1) were correctly identified less often than /ty/ syllables ($p < 0.01$).

Tests of simple main effects also indicated that the effect of speaker group was significant for both the /tu/ and

/ty/ syllables. *Post-hoc* tests (Tukey's HSD, $\alpha = 0.01$) revealed that the /tu/ syllables produced by the native French talkers (NF) were correctly identified more frequently than the /tu/ syllables produced by the experienced American talkers (NE-2) who, in turn, produced /tu/ syllables that were correctly identified more often than those of the inexperienced Americans (NE-1). *Post-hoc* tests revealed that for /ty/, on the other hand, syllables produced by the native French talkers (NF) were correctly identified more frequently than syllables produced by talkers in the two American speaker groups (NE-1, NE-2), but that there was no difference between the two American groups.

The mean percent correct identification scores for syllables produced by individual talkers in the three speaker groups are presented in Table I. In this table data have been collapsed across the seven subjects (i.e., listeners) and three speaking conditions. The data for individual talkers were analyzed in a mixed design analysis of variance in which speaker group was a between-group factor, and speaking task and syllable type were repeated measures.

In this analysis the effect of speaker group was again significant [$F(2,18) = 20.2, p < 0.001$]. *Post-hoc* tests revealed that the native French talkers (NF) produced syllables that were identified more correctly than those of the experienced Americans (NE-2) who, in turn, produced more identifiable syllables than the inexperienced Americans (NE-1) ($p < 0.01$).

The interaction between speaker group \times syllable type was not significant as it was in the "listener" analysis [$F(2,18) = 2.11, p = 0.145$]. Only six of seven native French talkers, and five of seven experienced Americans produced a more identifiable /tu/ than /ty/. Only six of the seven inexperienced Americans showed the opposite pattern, producing a more identifiable /ty/ than /tu/.

The effect of speaking task did not reach significance [$F(2,36) = 2.43$], but the interaction between speaking task and syllable type did [$F(2,36) = 9.04, p < 0.001$]. The syllable /tu/ tended to be correctly identified more often than /ty/ in the phrase and sentence tasks, whereas the reverse was true in the story task. However, tests of simple main effects indi-

TABLE I. The percentage of times /tu/ and /ty/ syllables produced by native speakers of French (NF), experienced American speakers of French (NE-2), and inexperienced American speakers of French (NE-1) were correctly identified. Each score is based on a total of 210 forced-choice identifications (7 listeners \times 2 replicate productions \times 3 speaking tasks \times 5 presentations).

| Talker | Speaker group | | | | | |
|-----------|---------------|--------|--------|--------|--------|--------|
| | NF | | NE-2 | | NE-1 | |
| | /tu/ | /ty/ | /tu/ | /ty/ | /tu/ | /ty/ |
| 1 | 99.5 | 95.1 | 92.3 | 82.8 | 40.3 | 61.3 |
| 2 | 99.5 | 91.5 | 89.0 | 66.1 | 89.5 | 3.8 |
| 3 | 98.5 | 97.5 | 35.8 | 84.1 | 41.2 | 94.1 |
| 4 | 97.1 | 99.0 | 99.0 | 93.8 | 24.3 | 79.0 |
| 5 | 99.0 | 89.5 | 80.7 | 30.1 | 23.6 | 69.0 |
| 6 | 94.7 | 90.5 | 99.0 | 46.5 | 44.5 | 97.2 |
| 7 | 98.0 | 92.6 | 84.3 | 97.0 | 50.2 | 87.1 |
| \bar{x} | 98.0 | 93.7 | 82.9 | 71.5 | 44.8 | 70.2 |
| s.d. | (1.70) | (3.61) | (21.9) | (25.1) | (22.1) | (32.0) |

cated that this interaction was of marginal importance. There was no significant effect of syllable type (/tu/ versus /ty/) in any of the three speaking tasks, and the effect of speaking task was not significant for either the /tu/ or the /ty/ syllables ($p < 0.01$).

C. Discussion

It is not surprising that listeners correctly identified more French syllables produced by French than American talkers. What is somewhat surprising is the extent to which correct identification rates differed between the experienced and inexperienced American groups. Listeners were able to correctly identify more of the /tu/ syllables produced by the experienced than inexperienced Americans, although there was no difference in the identifiability of /ty/ syllables produced by the two groups of Americans. This suggests that L2 learners' ability to produce a syllable containing a "new" vowel (i.e., /ty/) does not benefit from additional L2 experience, whereas the ability to produce a syllable with a vowel that has a counterpart in L1 (i.e., /tu/) does show an effect of additional experience.⁷

An examination of syllables produced by individual talkers did not support this conclusion regarding the effect of experience on the accuracy of /tu/ and /ty/ production. The interaction of speaker group \times syllable type did not reach significance in the "talker" analysis because 4 of 21 talkers failed to conform to the general pattern evident in the "listener" analysis. There was no overall difference in the correct identification of /tu/ and /ty/ because the experienced talkers tended to produce /tu/ better than /ty/, whereas the inexperienced talkers tended to produce /ty/ better than /tu/. Thus the data do not support the hypothesis that a new L2 vowel which has no direct counterpart in L1 (e.g., /y/) will be produced more successfully than a vowel which does have such a counterpart (e.g., /u/).

Only two of the seven experienced Americans, and two of the seven inexperienced Americans produced /ty/ syllables that were correctly identified 93% of the time, the mean for the native French speakers. Since the experienced Americans had been speaking French for about 10 years, this suggests that few American learners will match native French speakers in the ability to produce French /y/.

Their failure to do so may be the result of developing an incorrect articulatory strategy. Students in American schools are explicitly taught to produce French /y/ by placing the tongue in a configuration suitable for English /i/ and rounding the lips, as for English /u/. The experience of the first author in teaching beginning-level French classes is that this strategy results in a reasonable approximation to French /y/. Borden *et al.* (1981) noted that native English speakers unfamiliar with French were able to produce a recognizable /y/ on their first imitation trial.

If Americans produce French /y/ by recombining the articulatory features used for English vowels, it might explain their initial success in producing /y/, as well as their continued deviation from the phonetic norms of French. To produce French /y/ authentically, it may be necessary for the English learner to position the tongue differently for French /y/ than for English /i/. Delattre (1951) indicates

that French /y/ and /i/ are not distinguished primarily by lip rounding, but along a dimension he defines in terms of the anterior-posterior position of the tongue dorsum.

Supporting this are data presented by Linker (1982), which indicate little difference in upper and lower lip protrusion between French /i/ and /y/, little difference in the area of the mouth orifice, and little difference in the ratio of the horizontal to the vertical opening of the mouth during production of these vowels. Analogous data have been reported for Dutch vowels. Based on acoustic analysis, EMG data, and analysis by articulatory synthesis, Raphael *et al.* (1979) concluded that spectral differences between Dutch /i/ and /y/ were due to more than just a difference in lip rounding. It appeared that the maximum tongue constriction was somewhat greater and more anterior in the production of /i/ than /y/.

The use of lip rounding appropriate for English /u/ in producing French /y/ might also result in differences between English and French native speakers. According to Linker (1982), French /y/ is produced with somewhat greater protrusion of the upper and lower lips than English /u/, and has a substantially larger orifice size (3.53 sq cm compared to 0.23 sq cm). Since increasing the ratio of the mouth orifice relative to the length of the oral cavity importantly affects F_2 (Stevens and House, 1955), the use of lip rounding appropriate for English vowels in producing French /y/ may result in a difference between native and non-native speakers, even assuming that the tongue configuration is correct.

The inexperienced Americans' /u/ was misidentified as /y/ more than half the time. This is about what we would expect if they were producing French "tous" with an unmodified English /u/. Debrock and Forrez (1976) report average F_2 values of 987 and 2188 Hz for the /u/ and /y/ produced by monolingual French-speaking women. As part of a larger study (Flege, 1984c) the American talkers in this study produced English /u/ in a phonetic context comparable to the one in which French /u/ was produced here (i.e., in phrases like "two little boys"). The average frequency of F_2 in their English /u/ was intermediate to that for French /u/ and /y/ (1673 Hz).

The poor production of French /u/ by many of the inexperienced Americans might also have stemmed from a lack of awareness of the linguistic distinction between the /u/ and /y/ categories of French. If so, these talkers may have developed a perceptual target for French /u/ that embraced the /u/ and /y/ categories produced by native French speakers.

II. EXPERIMENT 2

Experiment 1 indicated that French /ty/ syllables produced by native speakers of American English were misidentified (as /tu/) about 30% of the time. The /tu/ syllables produced by experienced Americans were misidentified (as /ty/) an average of 17% of the time, as against 55% for Americans who were less experienced in French. This suggests that the relatively experienced Americans produced a perceptually more effective contrast between the French

vowels /y/ and /u/ than the less experienced talkers as the result of their greater experience. Experiment 2 directly assessed this vowel contrast using a paired-comparison task.

A. Methods

1. Subjects

Investigations of L1 acquisition suggest that adult listeners may overlook a distinction between two phones produced by children because of the tendency for speech to be perceived categorically (Monnin and Huntington, 1974; Macken and Barton, 1980; Maxwell and Weismer, 1982; cf. Locke, 1983). This can occur in instances where the child produces a reliable acoustic distinction between phones that represent a within-category phonetic difference for adult listeners.

The subjects chosen for this experiment were native English speakers. Native English speakers might be expected to be more sensitive to acoustic distinction(s) between /u/ and /y/ produced by other native English speakers than native speakers of French. There were six male and six female native English speakers with a mean age of 31 years. Each subject had studied French for at least 4 years in school. Five held advanced degrees in French and taught French, and nine had lived in a French-speaking environment for at least 3 months. Nine of the 12 had some training in phonetics, and all had normal hearing according to self-report.

2. Stimuli and procedures

The same 256 tokens of /ty/ and /tu/ used in experiment 1 were presented to subjects in a two-interval forced-choice task. Subjects were told they would hear one token of /tu/ and one token of /ty/ on each trial. Their task was to determine which member of the pair was most likely to be /ty/. No feedback, familiarization, or training was given.

The stimuli were blocked according to the speaking task (phrase, sentence, story) and phrase (#4 or #5) from which they had been edited. Within a block, each talker's production of /tu/ and /ty/ was presented four times, twice with intended /ty/ as the first member of the pair, and twice with /ty/ in the second position. This provided a total of 84 paired comparisons per block (21 talkers \times 4 presentations).

The six blocks of stimuli (2 replicate pairs of "tu/tous" \times 3 speaking tasks) were normalized for overall rms intensity and randomized separately for each subject. The order of blocks was counterbalanced across subjects. Stimulus presentation and response collection were run under the control of a laboratory computer (PDP 11/34). The interstimulus interval was set at 1 s. Presentation of each succeeding trial was triggered by the previous response, with a minimum intertrial interval of 1 s. The experiment lasted about 30 min, with a short break after the first three blocks.

The percentage of times the /ty/ syllables produced by each of the 21 talkers (3 groups \times 7 groups) was correctly chosen was computed. The maximum number of correct identifications for each of the three speaking tasks was 96 (2 "tous/tu" pairs \times 4 presentations \times 12 listeners). In addition, the percentage of times the 12 subjects (i.e., listeners) correctly chose intended /ty/ was calculated separately for

each of the three speaker groups (NF, NE-1, NE-2) in each of the three speaking tasks (phrase, sentence, story). These percent correct identification scores were based on a maximum of 56 possible correct judgments (2 "tu/tous" pairs \times 4 presentations \times 7 talkers).

B. Results and discussion

Table II presents the percentage of times the /ty/ syllables produced by the seven talkers in the three groups were correctly chosen. The data have been averaged over the three speaking tasks. The /ty/ syllables produced by the native French speakers (NF) were correctly chosen about 90% of the time. Listeners correctly chose 75% of the /ty/ syllables produced by the experienced Americans (NE-2), as against only 63% of the syllables produced by the inexperienced Americans (NE-1).

After arcsine transformation, the percent correct identification scores were submitted to a mixed design analysis of variance in which speaker group was a between-group factor and speaking task was a repeated measure. The effect of speaking task was not significant [$F(2,36) = 0.18$]. The effect of speaker group did reach significance [$F(2,18) = 6.65$, $p < 0.01$]. *Post-hoc* tests revealed that native speakers of French (NF) produced a more effective contrast between /tu/ and /ty/ than the experienced American talkers (NE-2) who, in turn, produced a better contrast than the inexperienced American talkers (NE-1) ($p < 0.01$). An examination of data for individual talkers indicated that listeners correctly chose /ty/ at better than chance levels for all seven native French talkers, but for only five of the seven experienced American talkers, and only three of the seven inexperienced Americans ($p < 0.01$ by the binomial probability test; Siegel, 1956).

Speaker group and speaking task served as repeated measures in a "listener" analysis of the same data. This analysis produced the same results as the "talker" analysis: no effect of speaking task, but a significant effect of speaker group [$F(2,22) = 41.33$, $p < 0.001$]. *Post-hoc* tests (Tukey's HSD, $\alpha = 0.01$) again indicated that the native speakers of French produced a more effective contrast between /tu/ and /ty/ than the experienced American talkers (NE-2) who, in turn, produced a more effective contrast than the inexperienced American talkers (NE-1).

These findings demonstrate the importance of experience for production of a potentially confusable pair of foreign language vowels such as /u/ and /y/. Our perceptual

TABLE II. The mean percentage of times the /ty/ syllables produced by native speakers of French (NF), experienced English speakers of French (NE-2), and inexperienced English speakers of French (NE-1) were correctly chosen in a paired-comparison task. An asterisk signifies that the /ty/ syllable produced by a talker was correctly chosen at a greater than chance rate ($p < 0.01$).

| Group | Talker | | | | | | | \bar{x} |
|-------|--------|-----|-----|-----|-----|-----|-----|-----------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
| NF | 87* | 96* | 90* | 95* | 90* | 90* | 88* | 90% |
| NE-2 | 90* | 73* | 52 | 90* | 47 | 83* | 95* | 75% |
| NE-1 | 52 | 46 | 82* | 55 | 52 | 85* | 70* | 63% |

evaluation indicated that although the experienced American talkers did not produce as effective a perceptual contrast between /y/ and /u/ as native speakers of French, they were nonetheless better at doing so than relatively less-experienced Americans.

This experiment also demonstrated that native speakers of English with some experience in French are able to effectively discriminate the /u/ and /y/ produced by native speakers of French. In experiment 1, native French-speaking subjects correctly identified the /y/ produced by native French speakers about 93% of the time. In this experiment, native English-speaking subjects correctly chose the /y/ produced by native French speakers about 90% of the time.

The American subjects' success in discriminating /u/ and /y/ does not necessarily demonstrate, however, that their perception of French /y/ and /u/ matches that of native speakers of French. It is possible they discriminated /y/ and /u/ on a purely auditory basis. It is also possible they were familiar enough with the phonetic properties of French /u/ and /y/ to discriminate these vowels phonetically, but without having precisely the same "prototype" or "perceptual target" as French native speakers.

III. ACOUSTIC ANALYSIS

The two perceptual experiments indicated that native speakers of English produced /tu/ and /ty/ less effectively than native speakers of French. In this section we acoustically examine the syllables presented in the perceptual experiments to determine how native and non-native speakers' production of those syllables may have differed.

Based on the perceptual results, we expected to find that the native French speakers produced a substantial spectral difference between /y/ and /u/, whereas the native English speakers—especially the inexperienced ones—produced a much smaller spectral distinction. The results of experiment 1 did not support the hypothesis that new phones (such as /y/) are produced more accurately than L2 phones with a close counterpart in L1 (such as /u/). This leads us to expect that, when measured acoustically, the difference between the French and American talkers will be the same for /y/ and /u/.

The perceptual experiments did not assess the accuracy with which French /t/ was produced. The acoustic analysis of /t/ in this section will permit us to test a hypothesis concerning why L2 learners ordinarily do not match native speakers of the foreign language being learned (Flege, 1981, 1984b). If L2 learners merge the phonetic properties of L1 and L2 phones judged to be equivalent (e.g., the /t/ of French and English) we expect to observe two phenomena. First, American talkers—even experienced ones—should produce French /t/ with VOT values that are longer than the short-lag values commonly observed for French monolinguals (about 20 ms, Caramazza and Yeni-Komshian, 1974), despite the fact that it may be less difficult to produce stops with short-lag than long-lag VOT values (Kewley-Port and Preston, 1974). Second, the French talkers should also produce French /t/ with longer VOT values than monolingual French speakers because of their massive exposure to the stops of English.

A. Methods

The 252 exemplars of /tu/ and /ty/ examined in experiments 1 and 2 were low-pass filtered at 4 kHz (Krohn-Hite model 3343) before being digitized at a sampling rate of 10 kHz with 12-bit amplitude resolution. As illustrated in Fig. 1, voice-onset time (VOT) was measured from the display of a graphics terminal (Tektronix model 4010) by setting a cursor at the beginning of the noise burst signaling stop release, and at the first upward-going zero crossing of the waveform, signaling onset of phonation.

The center frequencies of formants 1–3 in the vowels of /tu/ and /ty/ were estimated by means of linear predictive coding (LPC) analysis. Using an oscillographic display of the speech waveforms, a 256-point (25.6-ms) Hamming window was positioned so that its left tail coincided with the positive peak of the first pitch period in the approximately 50-ms periodic portion of syllables. Twelve linear prediction coefficients were calculated. Formant frequency values were then determined by picking amplitude peaks from the smoothed spectra using algorithms developed by Markel and Gray (1976).

The aperiodic portion (i.e., "VOT interval") of the syllables examined varied from about 30 to 80 ms. This means that some of the vowel measurements, especially those made of syllables with a very short VOT interval, may reflect the formant frequencies of consonant transitions into the "vowel" rather than just the "steady-state" portion of "vowels." However, since VOT did not differ across the three speaker groups (see Sec. III B), this should not invalidate between-group comparisons.

B. Results

1. VOT

Table III presents the VOT of /t/ in /ty/ and /tu/ syllables produced by talkers in the three speaker groups. These mean values represent the average of three speaking conditions and two replicate productions of both syllables. The VOT /tu/ and /ty/ was somewhat longer for the inexperienced Americans (63 ms) than for either the experienced Americans (50 ms) or native speakers of French (54 ms).

The VOT values measured in /tu/ and /ty/ syllables produced by each talker in the three speaking tasks were submitted to a mixed design analysis of variance in which speaker group (NF, NE-1, NE-2) was the between-group factor, and speaking task (phase, sentence, story) and syllable type (/tu/, /ty/) were repeated measures. There were no sig-

TABLE III. The mean duration, in ms, of the VOT in /tu/ and /ty/ syllables produced by native speakers of French (NF), experienced English speakers of French (NE-2), and inexperienced English speakers of French (NE-1). Each mean is based on 42 observations (7 talkers \times 2 replicate productions \times 3 speaking tasks); standard deviations are in parentheses.

| | /tu/ | | | /ty/ | | |
|------|------|------|------|------|------|------|
| | NF | NE-2 | NE-1 | NF | NE-2 | NE-1 |
| mean | 49.6 | 44.8 | 62.2 | 57.1 | 54.9 | 64.3 |
| s.d. | (14) | (13) | (15) | (17) | (23) | |

nificant interactions. Neither the effect of speaking task nor syllable type reached significance. The experienced Americans (NE-2) more nearly approximated the phonetic norms of French than the inexperienced Americans (NE-1). Although their VOT for /t/ averaged 13 ms shorter than that of the experienced Americans, the effect of speaker group also failed to reach significance ($p < 0.01$).

The data for individual subjects presented in Table IV are highly consistent with the grouped means presented in Table III. The most important point to note is that none of the talkers, including the native speakers of French, produced /t/ with an average VOT value of less than 35 ms. Thus, none of the talkers in this study closely resembled monolingual native speakers of French.

2. Vowel formant data

The mean frequency values of formants 1–3 in the /y/ and /u/ vowels are presented in Table V. These data have been averaged across the three speaking tasks and two replicate productions of each vowel by the seven talkers in each group. It is apparent that there was little difference between the groups for /y/. For /u/, there seems to be a between-group difference for F_3 , and an even larger difference for F_2 .

The frequency with which the 252 /ty/ and /tu/ syllables were identified as /ty/ in experiment 1 was correlated with the mean formant frequency values measured for those syllables. This analysis revealed that variations in F_2 accounted for 62% of the variance in the mean identification scores, as against only 9% for F_3 and 1% for F_1 . Therefore, only F_2 differences will be further discussed.

A closer examination of the F_2 data revealed clear differences between the American and French talkers, as well as a difference between the experienced and inexperienced Americans. When the 84 /u/ and /y/ vowels produced by talkers in each of three groups were plotted in an F_1 – F_2 space, there was practically no overlap in F_2 values between the /y/ and /u/ vowels produced by the native French speakers, some overlap for vowels produced by the experienced Americans, and almost complete overlap in F_2 for the /y/ and /u/ produced by the inexperienced Americans.

The F_2 difference between /u/ and /y/ averaged 715 Hz for vowels produced by the native speakers of French (NF). The F_2 difference between /y/ and /u/ was much

TABLE V. The mean frequency, in Hz, of formants 1–3 in the /u/ and /y/ vowels produced by native speakers of French (NF), experienced English speakers of French (NE-2), and inexperienced English speakers of French (NE-1). Each mean is based on 42 observations (7 talkers \times 2 replicate productions \times 3 speaking tasks); standard deviations are in parentheses.

| | /tu/ | | /ty/ | | | |
|-------|-----------|-----------|-----------|-----------|-----------|-----------|
| | NF | NE-2 | NE-1 | NF | NE-2 | NE-1 |
| F_1 | 283(45) | 262(22) | 266(40) | 260(35) | 247(24) | 265(51) |
| F_2 | 1387(211) | 1593(267) | 1909(193) | 2102(139) | 2006(297) | 2012(254) |
| F_3 | 2521(279) | 2624(186) | 2875(229) | 2779(387) | 2725(402) | 2840(212) |

smaller for the non-native speakers, averaging 413 Hz for the experienced Americans (NE-2), and only 103 Hz for the inexperienced Americans (NE-1).

A second finding was that the American talkers more closely matched the French talkers in producing /y/ than /u/. The formant values measured for /u/ and /y/ were submitted to a mixed design analysis of variance in which speaker group was the between-subjects factor, and vowel (/u/ or /y/) and speaking tasks (phrase, sentence, story) were repeated measures. There was a significant vowel \times speaker group interaction [$F = (2,18) = 11.84; p < 0.001$]. Tests of simple main effects revealed that the effect of speaker group was not significant for /y/ in any of the three speaking conditions, but that it was significant for /u/ in all three speaking tasks ($p < 0.01$). In each case, the French talkers produced /u/ with lower F_2 values than the experienced Americans who, in turn, produced /u/ with lower F_2 values than the inexperienced American speakers of French ($p > 0.01$). Tests of simple main effects also revealed that the French (NF) and experienced American talkers (NE-2) produced /y/ with significantly higher F_2 values than /u/ in all three speaking tasks, whereas the inexperienced Americans (NE-1) did not ($p < 0.01$).

As expected from the perceptual experiments, the effect of speaking task did not reach significance.

C. Discussion

1. VOT

The most striking characteristic of the VOT data is that the French talkers, who were proficient speakers of English, produced French /t/ with VOT values that substantially exceeded the approximately 20-ms VOT values commonly observed in the speech of French monolinguals (Caramazza and Yeni-Komshian, 1974). This confirms the prediction (Flege, 1981, 1984b) that L2 learning will affect the production of phones in L1. This prediction follows from the hypothesis that the perceptual target for an L1 phone—and eventually the motor plan used to realize it—changes as the foreign language learner is exposed to that phone's acoustically different counterpart in L2. More specifically, French speakers of English are hypothesized to merge the phonetic properties of French and English /t/ as the result of judging these acoustically different phones to be realizations of the same phonetic category (i.e., as the result of "interlingual identifications").

This hypothesis receives additional support from data reported by Caramazza *et al.* (1973). In that study, French

TABLE IV. The mean VOT, in ms, of the /t/ in /tu/ and /ty/ syllables produced by native speakers of French (NF), experienced English speakers of French (NE-2), and inexperienced English speakers of French (NE-1). Each mean is based on 12 observations (2 syllable types \times 2 replicate productions \times 3 speaking tasks).

| Talker | Speaker group | | |
|--------|---------------|-----------|-----------|
| | NF | NE-2 | NE-1 |
| 1 | 61.8 (17) | 38.8 (12) | 55.2 (11) |
| 2 | 62.3 (13) | 45.4 (12) | 49.1 (15) |
| 3 | 60.5 (7) | 40.9 (11) | 76.8 (10) |
| 4 | 52.1 (14) | 61.0 (16) | 46.4 (10) |
| 5 | 39.8 (10) | 55.2 (12) | 80.0 (27) |
| 6 | 57.0 (17) | 56.8 (23) | 57.4 (16) |
| 7 | 38.3 (6) | 47.9 (9) | 74.3 (10) |

speakers of English labeled stops in a VOT continuum differently than monolingual speakers of French. They also produced French stops with somewhat longer (English-like) VOT values than monolingual speakers of French. Unlike subjects in the Caramazza *et al.* (1973) study, the French talkers in this study all learned English as adults. At the time of the study they had lived for about 10 years in an English-speaking environment, and were using English as their primary language. This may explain why learning English seems to have affected our French talkers' production of French stops to a much greater extent than it did for French talkers in the Caramazza *et al.* (1973) study.

Another predicted effect of interlingual identification is that English learners of French will develop a perceptual target for French /t/ which merges the phonetic properties of French and English /t/, and that this will lead them, in turn, to "overshoot" the short-lag VOT values appropriate for French /t/. As predicted, the American talkers produced /t/ with substantially longer VOT values than monolingual French speakers. Not even those who were very experienced in French produced French /t/ with an average VOT value of less than 35 ms. Of the 168 American-produced stops examined, only seven had a VOT value of less than 30 ms (cf. Flege, 1980). This seems to confirm the prediction that, because of interlingual identification, adult learners of a foreign language will never succeed in producing L2 stops with complete accuracy when stops in their native language differ substantially in VOT from those in L2.⁸

2. Vowel formant data

The most important finding regarding vowel production was that the American talkers matched the French talkers in producing /y/ but not /u/. This supports the hypothesis that "new" L2 phones are produced more accurately than L2 phones which have a direct counterpart in the native language.

The French talkers produced /y/ with a mean $F2$ frequency of 2102 Hz. This represents a somewhat lower mean frequency than reported previously by Debrock and Forrez (1976) for five French monolingual women (2188 Hz). The experienced Americans produced /y/ with a mean $F2$ frequency of 2006 Hz, as against 2012 Hz for the inexperienced Americans. The small differences between the three groups in $F2$ for /y/ were not significant, although listeners correctly identified about 20% more of the /ty/ syllables produced by the French than American talkers in experiment 1. This suggests that the small between-group $F2$ differences we noted were perceptually relevant or, more likely, that some acoustic dimension(s) other than just $F2$ served to cue the identity of /y/.

The French talkers produced French /u/ with a substantially higher mean $F2$ value (1387 Hz) than previously reported for monolingual French speakers (987 Hz) by Debrock and Forrez (1976). This suggests that learning English influenced their production of French /u/, just as it influenced their production of French /t/. We hypothesize that the French speakers of English produced French /u/ with higher (more English-like) $F2$ values because they

judged the /u/ of English and French to be equivalent. By the same reasoning, the seeming lack of an L2 effect on /y/ may follow from their *not* judging French /y/ to be equivalent to an English vowel. This should be further tested in a study comparing the French vowel production of monolingual native speakers of French to that of French talkers who also speak English.

Many of our American talkers failed to produce a perceptually effective contrast between French /y/ and /u/ largely because they failed to accurately produce /u/. The French talkers produced /u/ with a mean $F2$ of 1387 Hz, compared to 1593 Hz for the experienced Americans, and 1909 Hz for the inexperienced Americans. The $F2$ value for French /u/ produced by the inexperienced Americans is somewhat surprising in view of the fact that adult L2 learners, at least those who are reasonably proficient in the foreign language, generally "approximate" the phonetic norms of a foreign language. In another study (Flege, 1984c) the inexperienced American talkers produced English /u/ with a mean $F2$ value of 1675 Hz. We would therefore have expected them to produce French /u/ with somewhat lower (more French-like) $F2$ values, rather than with $F2$ values that were actually higher (less French-like) than that of their English /u/.

An explanation for this finding is not immediately apparent from the data of this study. In experiment 2 we found that four of seven inexperienced Americans did not produce a perceptually reliable contrast between /u/ and /y/. The acoustic analysis revealed that this group of talkers did not produce a reliable $F2$ contrast between /u/ and /y/. One possibility is that at least some of the inexperienced Americans were not perceptually aware of the linguistic distinction between the French /u/ and /y/ categories.

Another possibility is that they were generally aware of the existence of this vowel distinction, but mistakenly thought that the word "tous" contains /y/ rather than /u/. A number of studies have shown that adult L2 learners sometimes replace an L2 phone which has a direct counterpart in L1 (e.g., /u/) with a new L2 phone (such as /y/) they have recently learned (see Flege, 1984b). This phenomenon may represent a form of overcompensation to the difficulty inherent in mastering the new sound system of a foreign language.

Both acoustic and perceptual criteria demonstrated that the experienced American talkers were aware of the linguistic distinction between French /y/ and /u/. Despite this, they produced French /u/ with a mean $F2$ frequency that was only slightly lower and thus more French-like (1593 Hz) than the $F2$ in their English /u/ (1670 Hz, Flege, 1984c). Thus even after many years of experience speaking French they seem to have done little to modify their production of /u/ in the direction of French phonetic norms.

IV. GENERAL DISCUSSION

The first conclusion to be drawn from this study is that adult native speakers of English may produce new phones in a foreign language (such as French /y/) more accurately than L2 phones which have a clear counterpart in the native language (such as French /u/). Listeners' identification of

vowels in experiment 1 revealed a tendency for inexperienced American speakers of French to produce the new vowel /y/ more accurately than /u/. This suggested that new L2 phones may be learned more rapidly than L2 phones which have a clear counterpart in L1. However, the reverse tendency was noted for more experienced American speakers of French and, as a result, the overall difference in the rate at which /u/ and /y/ were correctly identified was not significant.

An acoustic analysis of *F2* nonetheless revealed that /y/ was produced more accurately than /u/. The American talkers did not differ from the French talkers in producing the new vowel /y/, whereas they produced /u/ with significantly higher *F2* values than the French talkers. The experienced American talkers produced French /u/ with *F2* values equaling their French /y/, suggesting they may have confused these two vowel categories. The experienced American, on the other hand, produced French /u/ with *F2* values that closely corresponded to the *F2* measured in their production of English /u/ in another study.

The second conclusion to be drawn from this study is that experience enables adult learners of a foreign language to produce L2 phones with greater accuracy. Acoustic and perceptual analyses revealed that experienced American speakers of French produced a more effective contrast between French /u/ and /y/ than less experienced Americans in three different speaking tasks.⁹ They also produced French /t/ with somewhat shorter (more French-like) VOT values than the less-experienced Americans. These findings are consistent with the results of previous studies of foreign language speech production (e.g., Flege and Port, 1981; Port and Mitleb, 1980; Williams, 1980).

The observation that L2 learners sometimes approximate the phonetic norms of a foreign language leaves us with the intriguing question of why they seldom if ever match L2 native speakers in producing L2 phones that differ phonetically from their counterpart in L1. Many researchers (e.g., Scovel, 1969) have suggested that the ability of humans to learn new patterns of pronunciation diminishes near the end of childhood for neurophysiological reasons. We feel an alternate hypothesis worthy of further investigation is that previous phonetic experience impedes the formation of accurate perceptual targets for phones in L2 (Flege, 1981, 1984b). More specifically, we hypothesize that phones which closely resemble one another, such as the /t/ of French and English, mutually influence one another because language learners judge them to be acoustically different realizations of the same category.

This hypothesis is supported by the VOT data. The native French talkers in this study produced the /t/ in French words with VOT values that were intermediate to the short-lag and long-lag values typically observed for French and English, respectively. We hypothesize that the French talkers identified the prevocalic /t/ occurring in English words with the /t/ of French. We further hypothesize that, as a result, their perceptual target for French /t/ represented a merger of the phonetic properties of French and English /t/.

Similarly, we hypothesize that the American speakers

of French judged the /t/ of French and English as being different realizations of the same category. If the American talkers developed a perceptual target for French /t/ that merged the phonetic properties of French and English /t/, it means they were probably attempting to produce a stop with VOT values intermediate to those of monolingual speakers of French and English. We observed that although the Americans approximated the short-lag phonetic norm for French /t/, they—like the native French talkers—also produced French /t/ with VOT values that were intermediate to those of French and English monolinguals.

Previous research in L1 and L2 speech learning indicates that talkers' production of the VOT dimension eventually conforms to perception of the VOT dimension in stop consonants (see, e.g., Zlatin and Koenigsnekt, 1976; Williams, 1980). If accuracy in speech production is limited by the accuracy of the perceptual target that is developed during speech learning, native English speakers may never match native speakers of French unless they manage to develop two distinct perceptual targets, one for the /t/ of French and one for the /t/ of English. However, speech perception data reported by Caramazza *et al.* (1973) indicate that although French speakers of English labeled stops differently than French monolinguals, they did not label a French /t/ differently than an English /t/ (but cf. Elman *et al.*, 1977).

The hypothesis that English speakers will never produce /t/ with complete accuracy is further supported by the observation that even Americans talkers who held advanced degrees in French, had lived for more than a year in France, and had spoken French for more than 10 years did not produce French /t/ with the short-lag VOT values observed for monolingual speakers of French.

The acoustic measurements made of French vowels are also consistent with the hypothesized role of interlingual identification. We found that the Americans were able to accurately produce /y/ but not /u/. The "new" vowel /y/ does not have a direct counterpart in English, as does French /u/. As a result, it may not have been "identified" with any vowel in English, and thus escaped the limiting effect of previous phonetic experience.

The American talkers were much less successful in producing French /u/, a vowel which does have a clear counterpart in English. The inexperienced Americans seem to have produced French /u/ as if it were /y/. Their relatively poor production of /u/ may have resulted from a failure to perceptually differentiate the /y/ and /u/ vowel categories of French. The experienced Americans produced French /u/ much like the /u/ of English. This suggests that American learners of French may never accurately produce French /u/.¹⁰

This raises the question of why the American talkers seem to have approximated the VOT norm of French for /t/ but not /u/. We speculate that this is due to the nature of interlingual identification. There seems to be only a single phone in English (/t/) with which French /t/ will be identified. However, our acoustic evidence suggests that the inexperienced American talkers may have judged both the /u/ and /y/ of French to be equivalent to English /u/.

We alluded above to the possibility that the experienced

Americans may have produced the French word "tous" with the /u/ of English. An alternative hypothesis is that the perceptual target they developed for French /u/ represented a merger of the phonetic properties of the /u/ of French and English. Why their French /u/ much more closely resembled the /u/ of English than French, rather than falling at a point that is more nearly intermediate to the /u/ of French and English, is unclear. Perhaps their phonetic learning was slowed by the necessity of first distinguishing the /y/ and /u/ categories of French.

In summary, the results presented here indicate that adult learners of a foreign language do not always produce foreign language words with phones occurring in their native language. Both groups of American talkers produced the new vowel /y/ with relatively great accuracy. Both groups of Americans produced French /t/ with VOT values that were shorter than typical for English. This indicates that existing articulatory motor plans can be modified, and new ones established. Limits on the extent to which L2 learners approximate native speakers' pronunciation of an L2 phone which has an acoustically different counterpart in L1 may stem not from an inability to learn new forms of pronunciation, but from the interlingual identification of L1 and L2 phones. Judging acoustically different phones as belonging to the same phonetic category seems to underlie the process of speech perception. The continued operation of this perceptual process in L2 learning may lead to inaccurate perceptual targets for L2 phones which, in turn, limits the accuracy of L2 speech production.

We observed several differences between native and non-native speakers of French. The interpretation of these results was based on inferences concerning talkers' "perceptual targets" for L1 and L2 phones (see footnote 5). A great deal of further research is clearly needed to test these inferences. It will be important in future studies of L2 production to demonstrate which specific phones in the native and target language are judged to be equivalent (i.e., "identified" with one another), and to determine the extent to which the perceptual targets for L2 phones evolve as a function of experience with the foreign language.

¹The accuracy with which a learner produces the sounds of a foreign language can be objectively assessed in a variety of ways: (1) through the use of rating scale judgments by native speakers of the target language, (2) by calculating the frequency with which L2 phones are correctly identified, and (3) through acoustic analyses. This last method depends on a comparison of specific acoustic dimensions of an L2 phone produced by non-native speakers to the average value of that dimension in the speech of monolingual speakers of the target language.

²We examined the production of "tous" and "tu" in order to minimize the effect of differences in word familiarity. Studies of both L1 acquisition (e.g., Barton, 1980) and L2 learning (see Flege, 1984b) indicate that word familiarity may affect the extent to which phones are correctly produced. "Tu" and "tous" are among the first French words learned by non-native speakers owing to their high frequency of occurrence. "Tu" is the second person singular pronoun meaning "you"; "tous" is an adjective meaning "all."

³Cross-language auditory comparisons suggest that French /u/ is more "tense" or "peripheral" in the vowel space than its English counterpart (Delattre, 1953; Adamczewski and Keen, 1973; Valdman, 1976). If /u/ is articulated with a relatively more posterior tongue position in French than

English, one would expect it to be produced with lower F_2 values than English /u/ (see Lindblom and Sundberg, 1971). The comparison of French and English /u/ is complicated by the fact that English /u/ is produced more variably than its French counterpart, probably because there is no adjacent high vowel category in English (i.e., /y/) with which it risks being perceptually confused (Stevens, 1983). The results of several studies suggest that the second formant frequency (F_2) of /u/ is about the same (700–900 Hz) in French and English (Paterson and Barney, 1952; Delattre, 1951; Debrock and Forrez, 1976; Riordan, 1977). However, other studies indicate that English /u/ is produced with considerably higher (1000–1900 Hz) F_2 values than those reported for French /u/, especially in conversational speech (Stevens and House, 1963; Shockey, 1974; Labov, 1981). Thus if Americans produce French words with an English /u/, we expect them to produce French /u/ with higher F_2 values than monolingual speakers of French.

⁴If cross-language similarity judgments for vowels are based primarily on the position of the tongue and the resulting acoustic spectrum, it seems reasonable to think that listeners will judge French /y/ to be closer to English /i/ than to English /u/. If degree of lip rounding is important to similarity judgments, French /y/ might be judged to be closer to English /u/ than /i/. Americans seldom if ever realize French /y/ as an /i/-quality vowel (Walz, 1979), although speakers of certain West African languages are said to do so (N. Spector, 1983). Instead, they typically realize French /y/ as an /u/-quality (Walz, 1979), and at times produce French /u/ with an /y/-quality vowel (Gaudin, 1953). Evidence from speech production thus suggests that lip rounding might be more important to similarity judgments than the configuration of the tongue, Jakobson *et al.* (1952) note, on the other hand, that similarity judgments may depend as much on the *system* of phonological contrasts in the listener's native language as on the physical properties of phones found in L1 and L2. If so, no physical dimension, or combination of dimensions, may uniquely determine which L1 phone is judged to be "closest" to a new L2 phone.

⁵In this article we use the term "perceptual target" as a convenient cover term. Phoneticians have long debated what constitutes the "target" or "goal" for various phones, but what talkers aim to achieve in producing a phone is still unclear. It might be an "auditory" effect, the tactile and/or kinesthetic feedback associated with particular configurations of the speech articulators, or some combination of all three that varies according to phone or phone class. Research in recent years (e.g., Summerfield, 1979, 1983) suggests that a phonetic "target" or "goal" is not specified in terms of a modality-specific code, but is more abstract in nature. An alternative term we might have used here is "mental representation," for we conceive of a "perceptual target" as representing the talker's notion of how a phone "ought" to be produced. Another term we might have used is "prototype," for we consider the perceptual target to include all phonetic information, including language-specific and subcategorical information, pertinent to the production of a phone.

⁶Results reported by Elman *et al.* (1977) suggest that highly proficient bilinguals may have coexistent perceptual targets for counterpart phones in L1 and L2 (see Weinreich, 1953, 1963), but other previous studies suggest that bilinguals generally have a single perceptual target for counterpart phones in their two languages. This important issue invites further research.

⁷The dynamic spectral properties of the transition into the steady-state portion of syllables may have contributed to the identification of /tu/ and /ty/. However, the identifiability of the periodic portion ("vowel") in /tu/ and /ty/ syllables probably provides a good assessment of talkers' ability to produce /u/ and /y/. French- and English-speaking listeners are known to identify the vowels in French CV syllables (e.g., /tu/ and /ty/) as accurately as isolated French vowels (Gottfried, 1979, 1984).

⁸One reason for caution in accepting the conclusion that L2 learners never match native speakers of a target foreign language in producing stop consonants is that the American talkers in this study were not using French as their primary language at the time of the study. This conclusion should be further tested by examining the production of French by Americans who have spent a considerable period of time in a French-speaking environment and are using French as their primary language at the time of the study.

⁹Sociolinguists (e.g., Labov, 1972) have noted that variations in "attention to speech," as manipulated through the use of different speaking tasks, may affect native language speech production. Tasks that allow talkers to "pay attention" to their speech sometimes result in more "correct" productions of sounds (i.e., a more frequent production of variants found in the prestige dialect of the talkers' native language). In this study we observed that varying speaking task had no effect on L2 speech production. Acoustic and perceptual analyses revealed that American talkers pro-

duced the French syllables /tu/ and /ty/ with equal accuracy when reading a list of phrases, generating complete sentences from those phrases, and producing a series of sentences that were linked together in a spontaneous story. It seems reasonable to think that the manipulation of speaking tasks used in this study was sufficient to affect general vigilance or "attention to speech." We are forced to conclude that "attention to speech" has little effect on adults' production of L2 phones. We hasten to add, however, that no external measure was taken to demonstrate that general attention or vigilance changed across the three speaking tasks. It remains possible that "attention to speech" did *not* change as a function of speaking task, or that a manipulation of speaking task will influence the accuracy of L2 phones produced by learners who are not yet proficient in L2. Even the least experienced American talker in this study had spent nearly a year in France and was capable of the very demanding "story" speaking task in their foreign language. Weismer and Cariski (1983) suggest that the benefit of rehearsal for skilled motor control may be greatest in the early stages of acquisition, when cognitive factors are presumed to be relatively important. In the case of L2 learning, attention to speech might cease to influence production beyond the time learners establish a motor plan for producing L2 phones. Although somewhat uncertain, our negative findings in regard to the effect of speaking task is of some methodological importance for future studies of L2 speech production. It is simple to have talkers read phrase lists. Our results suggest that such a speech sample may provide a reasonable estimate of learners' production of L2 phones in other, more natural, speaking tasks.

¹⁰We can think of an important reason for tempering the conclusion that new L2 phones are produced better than L2 phones with cognates in L1. We considered the vowels in just two French words. The vowels in "tous" and "tu" might be unrepresentative of the way American talkers produce /u/ and /y/ in other words. The American talkers are likely to have learned "tous" and "tu" very early. As a result, these words might exemplify a nonoptimal approximation to the phonetic norms of French for /y/ and /u/ that remained "frozen" after later, more accurate, productions of these vowels were learned in other French words.

Adamczewski, H., and Keen, D. (1973). *Phonétique et Phonologie de l'Anglais Contemporain* (Armand Colin, Paris).

Barton, D. (1980). "Phonemic perception in children," *Child Phonology, Vol. 2, Perception*, edited by G. Yeni-Komshian, J. Kavanagh, and C. Ferguson (Academic, New York), pp. 97-116.

Bordon, G., Harris, K., Fitch, H., and Yoshioka, H. (1981). "Producing relatively unfamiliar speech gestures: A synthesis of perceptual targets and production rules," *Haskins Lab. Stat. Rep. Speech Res. SR-66*, 85-117.

Brière, E. (1966). "An investigation of phonological interference," *Lang.* 42, 769-796.

Caramazza, A., and Yeni-Komshian, G. (1974). "Voice onset time in two French dialects," *J. Phon.* 2, 239-245.

Caramazza, A., Yeni-Komshian, G., Zurif, E., and Carbone, E. (1973). "The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals," *J. Acoust. Soc. Am.* 54, 421-428.

Debrock, M., and Forrez, G. (1976). "Analyse mathématique des voyelles orales du Néerlandais and du Français, methode et résultats," *Rev. Phonétique Appl.* 37, 27-73.

Delattre, P. (1951). "The physiological interpretation of sound spectrograms," *Pub. Mod. Lang. Assoc.* 66, 864-875.

Delattre, P. (1953). "Les modes phonétiques du français," *French Rev.* 27, 59-63.

Elman, J., Diehl, R., and Buchwald, S. (1977). "Perceptual switching in bilinguals," *J. Acoust. Soc. Am.* 62, 971-974.

Elsendoorn, B. (1983). "Quality and quantity in English by Dutchmen: Two parameters inducing double Dutch," *Sound Structures, Studies for Antonie Cohen*, edited by M. van den Broecke, V. van Heuven, and W. Zonneveld (Foris, Dordrecht, The Netherlands), pp. 53-69.

Flege, J. (1980). "Phonetic approximation in second language acquisition," *Lang. Learn.* 30, 117-134.

Flege, J. (1981). "The phonological basis of foreign accent: A hypothesis," *TESOL Quart.* 15, 443-455.

Flege, J. (1984a). "The detection of French accent by American listeners," *J. Acoust. Soc. Am.* 76, 692-707.

Flege, J. (1984b). "The production and perception of foreign languages," in *Human Communication and Its Disorders, I*, edited by H. Winitz (Ablex,

Norwood), to appear.

Flege, J. (1984c). "The modification of phonetic representations during foreign language acquisition," in preparation.

Flege, J., and Hammond, R. (1982). "Mimicry of non-distinctive phonetic differences between language varieties," *Stud. Sec. Lang. Acquis.* 5, 1-17.

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

Gaudin, L. (1953). "Common mistakes in pronunciation," *French Rev.* 26, 451-460.

Gottfried, T. (1979). "Identification of French vowels," *Speech Communication Papers*, edited by J. Wolf and D. Klatt (Acoustical Society of America, New York), pp. 29-32.

Gottfried, T. (1984). "Effects of consonant context on the perception of French vowels," *J. Phon.*, to appear.

Jakobson, R., Fant, G., and Halle, M. (1952/1963). *Preliminaries to Speech Analysis* (MIT, Cambridge, MA).

Johansson, F. (1973). *Immigrant Swedish Phonology* (CWK Gleerup, Lund, Sweden).

Kewley-Port, D., and Preston, M. (1974). "Early apical stop production," *J. Phon.* 2, 195-210.

Kirk, R. (1968). *Experimental Design: Procedures for the Behavioral Sciences* (Brooks-Cole, Belmont, CA).

Koo, J. (1972). "Language universals and the acquisition of an unfamiliar sound," *Int. Rev. Appl. Linguist.* 10, 145-152.

Koutsoudas, A., and Koutsoudas, O. (1962). "A contrastive analysis of the segmental phones of Greek and English," *Lang. Learn.* 12, 211-230.

Labov, W. (1972). *Sociolinguistic Patterns* (University of Pennsylvania, Philadelphia).

Labov, W. (1981). "Resolving the neogrammarian controversy," *Lang.* 57, 267-309.

Lado, R. (1957). *Linguistics Across Cultures: Applied Linguistics for Language Teachers* (University of Michigan, Ann Arbor).

Le Bras, J. (1981). "Utilisation de la Reconnaissance Automatique de la Parole pour l'Apprentissage des Langues," unpublished Ph.D. thesis (Université de Rennes II, France).

Lindblom, B., and Sundberg, J. (1971). "Acoustical consequences of lip, tongue, jaw, and larynx movements," *J. Acoust. Soc. Am.* 50, 1166-1179.

Linker, W. (1982). "Articulatory and acoustic correlates of labial activity in vowels: A cross-language study," in *Working Papers in Phonetics*, Vol. 56 (Department of Linguistics, University of California, Los Angeles).

Locke, J. (1983). *Phonological Acquisition and Change* (Academic, New York).

Macken, M., and Barton, D. (1980). "The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants," *J. Child Lang.* 7, 41-74.

Marke, J., and Grey, A. (1976). *Linear Prediction of Speech* (Springer, New York).

Maxwell, E., and Weismer, G. (1982). "The contribution of phonological, acoustic, and perceptual techniques to the characterization of a misarticulating child's voice contrast for stops," *Appl. Psycholinguist.* 3, 29-43.

Monnin, L., and Huntington, D. (1974). "The relationship of articulatory defects to speech-sound identification," *J. Speech Hear. Res.* 7, 352-366.

Peterson, G., and Barney, H. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* 24, 175-184.

Politzer, R., and Weiss, L. (1969). "Development aspects of auditory discrimination, echo response, and recall," *Mod. Lang. J.* 53, 75-85.

Port, R., and Mittle, F. (1980). "Phonetic and phonological manifestations of the voicing contrast in Arabic-accented English," in *Research in Phonetics*, Vol. 1 (Department of Linguistics, Indiana University), pp. 137-166.

Raphael, L., Bell-Berti, F., Collier, R., and Baer, T. (1979). "Tongue position in rounded and unrounded front vowel pairs," *Lang. Speech* 22, 37-48.

Riordan, C. (1977). "Control of vocal-tract length in speech," *J. Acoust. Soc. Am.* 62, 998-1002.

Scovel, T. (1969). "Foreign accent, language acquisition, and cerebral dominance," *Lang. Learn.* 19, 245-254.

Shockey, L. (1974). "Phonetic and Phonological Properties of Connected Speech," in *Working Papers in Linguistics*, Vol. 17 (Department of Linguistics, Ohio State University).

Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, New York).

Spector, N. (1983). Personal communication.

Stevens, K. (1983). "Design features of speech sound systems," *The Production of Speech*, edited by P. MacNeilage (Springer, New York), pp. 247-

- Stevens, K., and House, A. (1955). "Development of a quantitative description of vowel articulation," *J. Acoust. Soc. Am.* 27, 484-493.
- Stevens, K., and House, A. (1963). "Perturbation of vowel articulations by consonantal context: An acoustical study," *J. Speech Hear. Res.* 6, 111-128.
- Summerfield, Q. (1979). "Use of visual information for phonetic perception," *Phonetica* 36, 314-331.
- Summerfield, Q. (1983). "Audio-visual speech perception, lip reading, and artificial stimulation," in *Hearing Science and Hearing Disorders*, edited by M. Lutman and M. Haggard (Academic, London), pp. 131-181.
- Valdman, A. (1976). *Introduction to French Phonology and Morphology* (Newbury House, Rowley).
- Walz, J. (1979). *The Early Acquisition of Second Language Phonology* (Helmholtz, Hamburg).
- Weinreich, U. (1953/1963). *Languages in Contact, Findings and Problems* (Mouton, The Hague).
- Weismer, G., and Cariski, D. (1983). "On speakers' ability to control speech mechanism output: Theoretical and clinical implications," *Speech and Language: Advances in Basic Research and Practice, Vol. 10*, edited by N. Lass (Academic, New York).
- Williams, L. (1980). "Phonetic variations as a function of second-language learning," *Child Phonology, Vol. 2, Perception*, edited by G. Yeni-Komshian, J. Kavanagh, and C. Ferguson (Academic, New York), pp. 185-216.
- Zlatin, M., and Koenigsknect, R. (1976). "Development of the voicing contrast: A comparison of voice onset time in stop perception and production," *J. Speech Hear. Res.* 19, 93-111.