



# Canadians in Alabama: a perceptual study of dialect acquisition in adults

**Murray J. Munro\***

*Department of Linguistics, Simon Fraser University, Burnaby, BC, Canada V5A 1S6*

**Tracey M. Derwing**

*University of Alberta, Edmonton, AB, Canada*

**James E. Flege**

*University of Alabama at Birmingham, Birmingham, AL, U.S.A.*

*Received 24th May 1998, and accepted 2nd December 1999*

---

Anecdotal evidence suggests that adults who move from one dialect area to another begin to acquire some of the speech characteristics of the second dialect (or “D2”). We put this observation to an empirical test in the present study by examining the speech of Canadians who have moved to Birmingham, Alabama. Speech samples produced by these people, as well as by native Alabamans and Canadians who had not emigrated were rated by two groups of listeners. This form of evaluation allowed us to determine whether dialect acquisition, if any, was perceptible. The fact that listeners from both Canada and Alabama reliably rated the speech of Canadians who had moved to Alabama as exhibiting an intermediate degree of “American accent” indicates that adults can indeed acquire the phonetic aspects of a D2 to the extent that listeners from both the first dialect and D2 areas can distinguish them from other native speakers of either dialect. A follow-up analysis allowed us to pinpoint some specific phonetic properties of the migrant group’s speech that may have played a role in the listeners’ judgments. In several respects these findings parallel the results of studies of second-language phonetic learning in adults.

© 1999 Academic Press

---

## 1. Introduction

There are reasons to suppose that, in some respects, learning a second dialect (D2) parallels learning a second language (L2). For the learner, both processes may entail acquiring new lexical items and new syntactic patterns, as well as making modifications to the phonetic repertoire. A learner’s acquisition of aspects of a D2 may well be evident to friends and relatives who speak the first dialect (D1). In fact, anecdotal reports of migrants who are said to sound like speakers of D2 are fairly common.

\*Correspondence to M. J. Munro. E-mail: [mjmunro@sfu.ca](mailto:mjmunro@sfu.ca)

However, such anecdotal reports cannot in themselves be regarded as convincing perceptual evidence of second dialect learning. One obvious reason is that an interlocutor may incorrectly attribute idiosyncrasies in a speaker's lexicon, syntax, or pronunciation to the speaker's having lived in another region. Another reason is that the judgment that a speaker sounds different from oneself may be made *post hoc*. An interlocutor who knows that an acquaintance has lived in another dialect area may be primed to imagine that the acquaintance sounds "different". Another question that has not been addressed is whether speakers of the D2 are able to recognize dialectal change in D1 speakers who have resided in the D2 area.

In this study, we used an experimental approach common in L2 speech research to explore perceptions of D2 speech learning in adult Canadians living in the southern United States. Our procedure entailed the collection of accent ratings from untrained listeners representing both D1 and D2 who were blinded to the speakers' place of origin. If the listeners reliably assigned different ratings to Canadians who had remained in Canada and Canadians who had lived in the southern U.S.A., we could infer that the basis for their ratings had something to do with changes in the speech productions of the latter group.

The notion that speakers are sometimes capable of acquiring a second dialect has been substantiated by empirical studies of migrant children. In his research on Canadian immigrants in the U.K., for example, Chambers (1992) found evidence that lexical changes are an important aspect of the early stages of D2 acquisition in children. Moreover, Payne (1980) documented the acquisition of specific phonological variables by children of various ages.

However, the published research on dialect acquisition has some notable limitations. The first is that most D2 studies focus on children. In fact, we know of little experimental work examining dialect learning in adults. Yet, the degree to which adult learners are capable of acquiring a D2, and the extent to which listeners recognize such acquisition, are issues of considerable theoretical importance. With respect to speech learning across the life span, an extensive body of evidence supports the notion that late learning (i.e., learning after early childhood) of an L2 is associated with increasingly less native-like pronunciation (Asher & Garcia, 1969; Oyama, 1976; Patkowski, 1990; Tahta, Wood & Lowenthal, 1981). In fact, a strong linear relationship has been observed between the age at which L2 acquisition begins and listeners' ratings of the degree of foreign accent that a speaker exhibits (Flege, Munro & MacKay, 1995). This relationship may sometimes be the result of older learners' failure to establish new phonetic categories for vowels and consonants in their L2 or their failure to make appropriate within-category adjustments (Flege & Eefting, 1987; Flege et al., 1995; Munro, Flege & MacKay, 1996; Flege, 1997). Such tendencies are often taken as evidence in favor of a sensitive period for language acquisition (Scovel, 1989; Long, 1990).

The evidence that native-like speech production is highly unlikely in late second-language learners has figured prominently in current theories of second-language acquisition. For instance, Pulvermüller & Schumann (1994) have proposed a two-factor model that attempts to account for variable success in L2 acquisition in terms of biological factors (i.e., a sensitive period) and motivational factors. If these influences are indeed operative in L2 learning, they might apply to D2 learning as well. Nevertheless, according to Flege's speech learning model (Flege, 1995), the production and perception of speech sounds remain subject to adaptation across the life span. If so, we should expect

adult migrants to a new dialect area to be able to acquire some but not all of the features of the D2.

In fact, the available evidence on biological factors suggests that late learners of a D2 will probably not learn to produce all the characteristics of the D2 in a completely native-like fashion — that is, to speak it exactly like individuals who learned it from birth. Payne's (1980) data, for instance, indicated that the earlier in life children were first exposed to a D2, the more likely they were to successfully acquire specific D2 phonetic variables. Other research suggests that the effect of late D2 learning may be less perceptible in social interactions than early D2 learning. In the Krashen and Seliger (1975) study of migrants who had lived in New York City for at least two years, nearly 60% of those who moved between the ages of 3 and 9 said that they had been taken for New Yorkers when travelling out of state. In contrast, only about 19% of those arriving in the city after age 15 reported such an experience. It is quite likely, as well, however, that the basis for their interlocutors' impressions was not only speech characteristics. Chambers (1992), in fact, concluded that the acquisition of pronunciation features by D2 learners tends to lag considerably behind the acquisition of lexical replacements and that complex phonological phenomena are much less likely to be acquired by older migrants than by younger ones.

Empirical research has linked motivation factors to success in acquiring an L2 accent (Moyer, 1999). Here one finds interesting differences between learning a second language and learning a different dialect of one's own native language. Someone may be motivated to acquire an L2 for the purpose of communication in a new cultural environment. However, if two dialects are mutually intelligible, as is the case with Canadian and southern US English, there may be little need for D1 speakers to make any modifications to their speech in order to be understood. Moreover, it seems reasonable to think that at least some of the Canadians who participated in the present study may have been fairly resistant to adopting southern American speech patterns because they may have associated them with low-prestige<sup>1</sup> English (see Lippi-Green, 1997). On the other hand, if Canadian immigrants to the US wish to be accepted and seen as members of their new community, they may be motivated to acquire at least some aspects of D2.

Discussions of motivational issues often assume some degree of volitional control on the part of learners. To varying degrees, L2 learners may consciously strive to pronounce their new language in a native-like way. However, L2 speech learning may also entail perceptually guided production changes of which the learner has little or no awareness. Sancier & Fowler (1997), for instance, have examined the phenomenon of gestural drift in L2 acquisition. In their research and in the work by Major (1992), L1 speech has been shown to undergo phonetic change (specifically in voice onset time) as a result of L2 learning. Moreover, such phonetic change has been found to be noticeable to other L1 listeners (Sancier & Fowler, 1997). The findings of Sancier & Fowler are reminiscent of the Giles & Smith (1979) notion of convergence, a "process by which individuals shift their speech styles to become more like that of those with whom they are interacting" (p. 46).

Another limitation of much previous work examining dialect learning is that much of this work has focussed on a relatively narrow range of phenomena. The variables of

<sup>1</sup>Canadians sometimes have stereotyped notions about Southern Americans based on negative depictions in the media. In an informal study by the first author of Canadian undergraduate students' reactions to the Alabama speakers in this experiment, it was found that they tended to rate Alabama speakers as less educated and more racist than Canadian speakers.

interest are usually particular vowel and consonant segments that vary across dialects and are thought likely to stratify according to factors such as age, gender, and social class. Thus, researchers may consider a particular set of phones or putative phonological rules<sup>2</sup> and explore, through interviews and other speech elicitation techniques, the extent to which learners acquire them. The results of such studies may, at least in certain instances, be affected by experimenter bias. This is because the evaluation is usually made by one researcher who knows in advance the background of the speakers, as in the previously mentioned work by Chambers (1992) and Payne (1980). Moreover, it remains unclear how the speech of D2 learners might be perceived by linguistically untrained, disinterested listeners from D1 or D2 background. It is conceivable, for example, that some D2 learners may make phonetic changes that can be transcribed by trained observers and that are salient to listeners from the D1, but that are not readily noticed by listeners from the D2. This would parallel the Sancier and Fowler (1997) findings for gestural drift in L1. Still another possibility is that migrants to the D2 area may not acquire the specific phenomena typically considered by linguists, yet still be perceived as having made dialectal adjustments. This may arise if the acquired properties of D2 speech are not easily characterized in terms of segments, phonetic features, or phonological rules, as may be the case with some prosodic and voice quality phenomena.

Obviously, the degree to which speakers are *perceived* to have acquired D2 may be an important factor in how speakers from both the D1 and D2 communities respond to them and in how they are accepted in either community. We propose, then, that data in the form of holistic listener ratings of D2 speech will provide a useful and necessary complement to the more narrowly focused data on dialect learning that are currently available.

In this study, we assessed perceived dialectal change by collecting accent judgments from native speakers of Canadian and Alabama English who listened to utterances produced by three groups of speakers: Canadian English speakers living in Canada, native speakers of Alabama English living in Alabama, and Canadian immigrants to Alabama. One of the most obvious ways in which the Canadian and Alabama dialects differ phonetically is in terms of vowel properties. Canadian English typically exhibits the Canadian Raising phenomenon (Chambers, 1973) according to which the vowel nuclei in the diphthongs [aj] and [aw] are “raised” to [ʌj] and [ʌw] before voiceless consonants. Alabama English on the other hand, shows no such pattern; in fact, one salient characteristic of this dialect is the use of monophthongal [a] where Canadians would use [aj] in at least some contexts (see Labov & Ash, 1997). Thus, in addition to examining the listener ratings mentioned earlier, we tested whether adoption of the Alabama vowels can be noticed in the migrant Canadian adults.

Previous research has shown that native listeners are very sensitive to L2 speakers’ deviations from native speaker norms (Flege, 1984) and that even phonetically untrained listeners can readily rate L2 speech samples on accentedness scales. Such judgments have been shown to be reliable (Ryan, Carranza & Moffie, 1977; Brennan & Brennan, 1981; Thompson, 1991) and related to factors such as age of L2 learning and length of residence in the L2 area (Flege & Fletcher, 1992; Flege et al., 1995), as well as to independently measured properties of the stimuli such as phonetic error counts and acoustic features

<sup>2</sup>Here we set aside the debate over whether language (or dialect) acquisition entails learning phonological rules at all, a proposal that has come under serious criticism in connectionist accounts of L1 and L2 acquisition (see, e.g., Pulvermüller, 1995).

(Anderson-Hsieh, Johnson & Koehler, 1992; Munro & Derwing, 1995; Derwing & Munro, 1997). Although some research has shown that listeners from different dialect areas can detect L2 accent in immigrants to one of the dialect areas (e.g., Flege, Frieda & Nozawa, 1997), this approach has not been previously used in D2 acquisition research.

## 2. Experiment I

### 2.1. Method

#### 2.1.1. Speakers

Three groups of speakers were recruited, each consisting of five females and five males between 20 and 46 years of age. All were white middle class individuals with normal hearing, according to self-report. The CC (Canadians in Canada) group had at least an undergraduate degree, were born and raised in Canada, and were living in Edmonton at the time of the study. The second group, Group CA (Canadians in Alabama), consisted of speakers of Canadian English who had taken up residence in Birmingham after living in Canada until at least the age of 18 years. We used this minimum age of arrival in the U.S. as a criterion to ensure that all participants would have passed the hypothesized sensitive period for language acquisition (Scovel, 1989; Long, 1990). Most were health-care professionals with at least an undergraduate degree who had regular contact with the public and who were likely to interact on a regular basis with native speakers of Alabama English. The mean length of time spent in the U.S. was 7.7 years, with a range of 1–23 years. Five of the participants from this group had lived in other U.S. cities besides Birmingham (in Illinois, Indiana, Missouri, Oregon, and Tennessee). The speakers from both the CC and CA groups were raised speaking Inland Urban Canadian English. Chambers (1991) has stated that “all of urban middle-class Canada (except Québec and Newfoundland) speaks a remarkably homogeneous accent” (p. 91). In addition, the CC and CA groups were very comparable in terms of education level and social status. There was no reason to expect any systematic difference in speech patterns between them. The members of the third group (AA) had grown up in Northern Alabama. All were either graduates of or students at the University of Alabama at Birmingham.

#### 2.1.2. Procedure

Extemporaneous speech samples were obtained from each of the 30 speakers. Recordings of the CC group were made using a TEAC V-437C cassette recorder in the Department of Linguistics at the University of Alberta. Recordings of the AA and CA groups were made using a Sony Pro-II tape recorder in the Department of Rehabilitation Sciences at the University of Alabama at Birmingham. The speakers were seated in a sound-treated booth and were shown a cartoon story (with no accompanying text).

The participants were asked to describe, in narrative fashion, the events depicted in the cartoon (two men set out on a hunting trip, but, after a series of misfortunes, end up taking photographs instead of hunting deer). The speech samples were elicited in the context of foreign-accent studies being carried out by the authors. Prior to the recording session, each group had rated a variety of utterances for degree of non-native accent. The participants were then asked whether they would mind providing us with a short speech sample. They were not specifically told until *after* the speech sample was collected that

we were interested in the “Americanness” of their dialect. The recording was done by the first author (MJM). The fact that the Canadians’ speech samples were elicited by a fellow Canadian was thought likely to work against finding between-group differences. To the extent that the Canadians accommodated their speech to that of their Canadian interlocutor, the results obtained here may have underestimated the actual effect size.

The speech samples were digitized and stored on disk as audio files so that they could be presented in random order to the listeners. From each narrative, a 10 s extract was digitized at 22 kHz with 16-bit resolution using a Macintosh computer. The extracts were all taken from the beginning of each narrative to ensure similar content.

The 30 samples were rerecorded on a stimulus tape, twice each in random order. In addition, eight samples were randomly selected for presentation at the beginning of the tape as warm-up items. A 2 s pause was recorded after each sample so that the listeners would have sufficient time to record responses on an answer sheet. The number of each item was taped so that the listeners could stay in step with the recording by noting numbers on the response sheets.

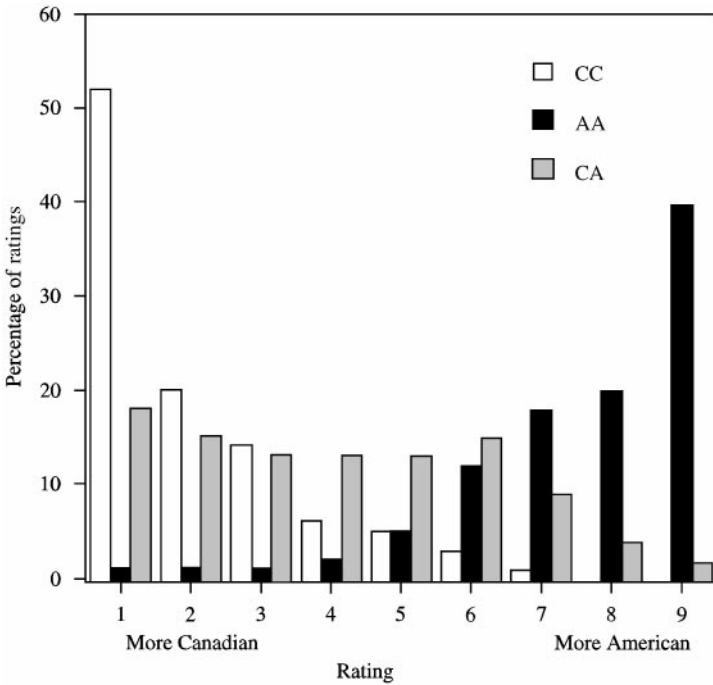
### 2.1.3. Listeners

A group of 22 native Canadian residents of Edmonton, all of whom reported normal hearing, auditorily evaluated the speech samples just described. The ratings were collected at the University of Alberta during one of two listening sessions. The listeners, who ranged in age from 19 to 45 years ( $M = 31$ ), were told that they would hear a number of speech samples from talkers who were Canadians, Alabamans, and Canadians living in Alabama. They were asked to rate the speech excerpts for degree of American accent using a scale that ranged from “very Canadian” (1) to “very American” (9). Four excerpts (two each from a Canadian and an Alabaman, neither of whose voices appeared in the experimental stimuli) were then played as examples of the kinds of utterances that would be presented during the test. The listeners were given no information about where the individual talkers were from. The stimulus tape was then played and the listeners assigned their judgements to the eight warm-up items, and then to the rest of the stimuli. In all, 1320 ratings were collected.

## 2.2. Results

We assessed intra-rater reliability by computing Pearson correlations ( $r$ ) for each listener between the ratings on the first presentation of the stimuli and the ratings on the second presentation. In general, the reliability was high, with only three of the listeners showing correlations of less than 0.80. We estimated inter-rater reliability using the method recommended by Hatch and Lazaraton (1991), which entails transforming all intercorrelations to Fisher Z scores. The reliability coefficient ( $r$ ) of 0.77 indicated a satisfactory level of inter-rater agreement, given that the listeners had received no training in the task.

Fig. 1 is a histogram showing the frequencies of the ratings (1 to 9) given by the Canadian listeners to the three groups of talkers. The ratings of the CC group (plain bars) tended to cluster near the left side of the scale. In fact, more than half were ratings of 1. Very few ratings higher than 4 were assigned to this group, and not one of the ratings was higher than 7. The dark bars show the frequencies of the ratings assigned to the AA group. This time the modal rating category was 9, and very few ratings lower than 6 were assigned. The ratings assigned to the immigrant Canadian (CA) group are represented by



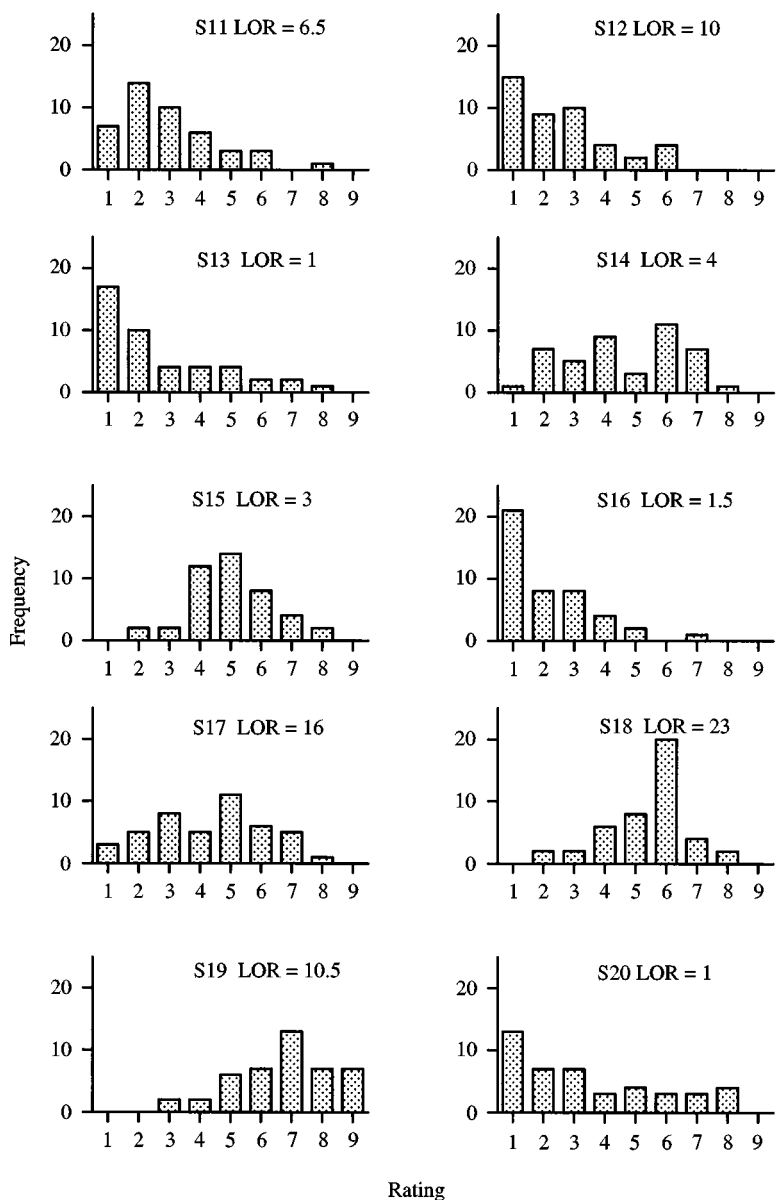
**Figure 1.** Frequencies of each rating (from 1 to 9) assigned by the Canadian listeners to the samples from each of the three groups of speakers. □, CC; ■, AA; ▒, CA.

the grey bars. While some ratings of 1 were assigned, overall the ratings tended to fall in the center of the range. In summary, the speech samples produced by the Canadians living in Alabama received ratings that were generally intermediate to those obtained for the Alabamans and the Canadians recorded in Canada.

The mean ratings assigned to each talker group (pooled over talkers) were 2.1 for the Canadians in Edmonton and 7.6 for the native Alabamans, with an intermediate value of 4.0 for the immigrant Canadians. A repeated measures ANOVA on the mean scores assigned by the listeners to each of the three groups revealed a significant effect of speaker group,  $F(2, 42) = 514.46$ ,  $p < 0.01$ , and *post hoc* Tukey tests indicated that all three groups differed significantly from each other at the 0.05 level. In short, the CC group was rated as most Canadian-sounding and the AA group was rated as most American-sounding. The CA group received an intermediate rating that differed significantly from the ratings assigned to both the other groups.

These untrained listeners assigned scores to the talkers that fell into three distinct categories. When separate ANOVAs were performed on individual listener data, they all revealed significant effects of talker group on the ratings, even when a very conservative  $p$  level of 0.002 was used to control for experiment-wise error. Furthermore, every listener showed the same ordering of mean ratings for the three talker groups.

We also examined the data for the individual talkers (pooling across listeners). A one-way ANOVA revealed a significant effect of speaker group,  $F(2, 27) = 61.44$ ,  $p < 0.01$ ,



**Figure 2.** Distributions of Canadian listeners' ratings for the individual speakers in the Canadian immigrant group.

and once again, Tukey *post hoc* tests indicated that the mean score assigned to every group differed significantly from the mean for every other group ( $p < 0.05$ ).

Some differences were observed in the patterns of ratings assigned to individual speakers in the CA group. Fig. 2 shows individual histograms for each of the 10 talkers. Some of these speakers (for example, number 13 and number 16) received ratings that clustered towards the “Canadian” end of the scale; some received ratings more in the



middle (for example, talker 15), and two (talkers 18 and 19) received a fairly large number of ratings toward the "American" side of the scale. Talker 19 actually received a mean rating (6.7) that was higher than or equal to the ratings assigned to three speakers from the native Birmingham group.

Although there is some indication that the amount of time spent in the US may partially predict these differences, this data set does not permit formal analysis of the effect of length of residence (LOR) in the U.S. (Fig. 2 also shows LOR in the U.S. for each speaker). The speakers who had lived in the U.S. for 10.5 and 23 years (talkers 19 and 18, respectively) received the highest mean ratings. In contrast, the talker with an LOR of 1.5 years received the lowest rating. It seems clear, however, that differences in length of residence cannot fully explain the differences in ratings seen here. For example, talker 12 had lived in the U.S. for 10 years, yet received a mean rating of only 2.6.

A potential problem with the results reported above is that factors other than the phonetic properties of the speakers' utterances may have contributed to the listeners' ratings. Canadians, for instance, might readily recognize a learner of an American D2 because of simple differences in the pronunciation of specific individual words, such as the frequently heard American /*ɹ*awt/ ('route') in place of Canadians' virtually exclusive use of /*ɹ*ut/. Another possible identifying feature might be the use of conspicuous new lexical items (e.g., some Americans use the word "soda" while Canadians almost universally say "pop"). Our preliminary examination of the narratives suggested that phenomena such as these were unlikely to have been a factor in the results. However, to rule out the possibility of lexical cues, we transcribed each of the narratives in standard orthography and presented the written versions to seven Canadian raters (who had not participated in the listening task); they used a 9-point scale to rate the passages for degree of "Americanness". One-way ANOVAs performed on the raters' data failed to reveal significant effects of speaker group for any of the listeners ( $p > 0.10$  in all cases). There was no evidence, then, that the lexical or syntactic properties of the utterances provided reliable cues to the speakers' dialects.

### 3. Experiment II

The results of Experiment 1 indicate that many of the speakers in the Canadian immigrant group had acquired aspects of the D2 and that, in one instance, this acquisition had made the speaker indistinguishable (from the perspective of Canadian listeners) from native speakers of Alabaman English. We carried out Experiment 2 to determine whether Alabama listeners would have similar impressions of the speakers.

As discussed by Flege (1984), there are two ways that listeners might detect foreign accent, or, by extension, dialectal differences. They might detect divergences from the norms of their native dialect, or they might have stored in long-term memory the properties of some other variety of speech and detect those properties in the speech sample. Flege & Hammond (1982), for example, showed that American college students from Florida who had been exposed to Spanish-accented English were able to produce English sentences with characteristics that are actually present in Spanish-accented English.

The listeners in Experiment 1 might have used either basis in their judgments of the Americanness of the CA group because they had almost certainly been exposed to Southern varieties of American English through television and movies. Alabama

listeners, however, are unlikely to have had such exposure to Canadian English, because the Canadian variety is rarely identified in the American media. Therefore, if Alabama listeners were found to be able to distinguish the CA group from the CC group, it could not be due to their knowledge of the divergences from the norms of Canadian English. Rather, it could only be ascribed to their ability to gauge degrees of divergence from their native dialect of English.

### 3.1. Method

#### 3.1.1. Speakers

The same stimulus items used in Experiment 1 were used in this experiment.

#### 3.1.2. Listeners

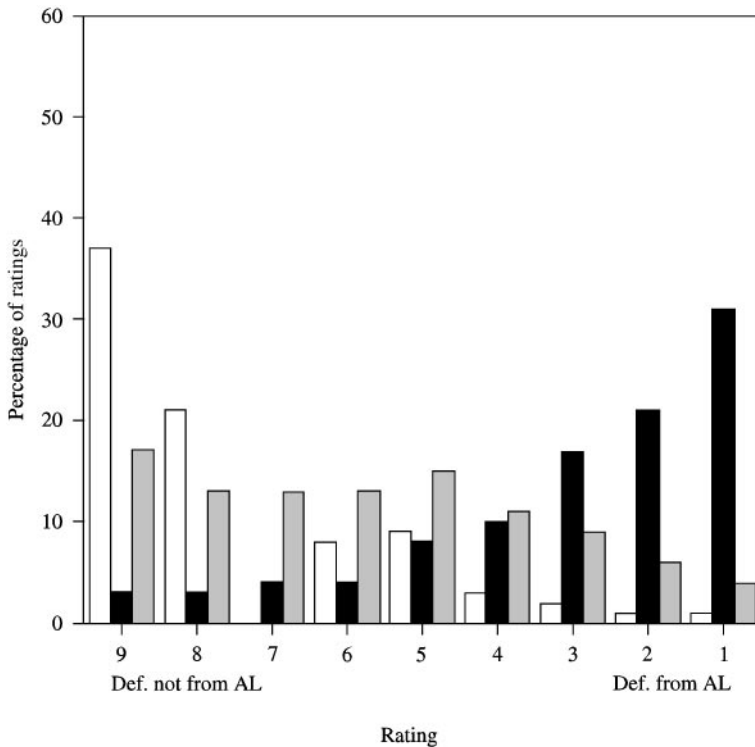
For the new listening task we recruited 27 undergraduate students studying psychology at the University of Alabama at Birmingham. All of them were born and raised in Alabama and were residing in Birmingham at the time of the study. None of them had ever lived in Canada. They had an average age of 21 years (range: 19–25). All reported having normal hearing except one listener whose data were excluded in the analyses reported below.

#### 3.1.3. Procedure

The Alabama listener data were collected at the University of Alabama at Birmingham, using an almost identical procedure to the one described in Experiment I. The only difference was in the wording of the rating scale, in that “1” signified “definitely from Alabama” and “9” meant “definitely not from Alabama.” The rationale for the different labels on the rating scales from Experiment 1 to Experiment 2 was a pragmatic one. One of the difficulties in dealing with the two distinct dialect groups was finding a common ground for describing the same set of stimuli. In order to capture the listeners’ ability to distinguish between D1, D2, and an intermediate variety, given that our two listener groups had limited exposure to each other’s dialects, we had to identify labels that they could easily understand. Because the Canadian listeners were unlikely to be able to distinguish between accents from different parts of the American South, we decided to label the scale they used with the terms “Canadian” and “American” at the extremes. The Alabama listeners, on the other hand, were unlikely to have a sense of a Canadian accent. We therefore determined that a contrast between “definitely from Alabama” and “definitely not from Alabama” would allow us to access their intuitions most successfully. The listeners were told that this was a study designed to “help us understand more about how adults acquire new speech patterns”. They were informed that they would hear speech from people who had always lived in Alabama, from people who had never lived in Alabama, and from people who had moved to Alabama from other states or countries.

### 3.2. Results

Intra-rater reliability was assessed in the same manner as in Experiment 1. Three of the listeners had  $r$  values much lower than those of the Canadian English listeners (i.e., less



**Figure 3.** Frequencies of each rating (from 1 to 9) assigned by the Alabamian listeners to the samples from each of the three groups of speakers. The horizontal axis has been reversed to permit easy comparison with Fig. 1. □, CC; ■, AA; ▒, CA.

than 40% of the variance in the second set of ratings could be accounted for by the ratings on the first set). We decided to exclude these listeners because of their lack of consistency.<sup>3</sup> Inter-rater reliability for the remaining 23 listeners was calculated using the same method as in Experiment 1. The reliability coefficient of 0.76 was nearly identical to that observed for the Canadian listeners.

Fig. 3 shows the frequencies of the Alabama listeners' ratings from 9 to 1 assigned to the samples from each of the three speaker groups. Note that we have reversed the horizontal axis on this figure to permit direct comparison with Fig. 1 for the Canadian listeners. There are some clear similarities between the judgements assigned by the two groups of raters. Parallel to the case in Experiment 1, the Alabama listeners overwhelmingly rated the AA group toward the Alabama end of the scale and the CC group toward the non-Alabama end. Once again, the CA group showed an intermediate distribution of scores. The mean ratings (pooled over talkers) were 7.4 for the CC group and 3.0 for the AA group, with an intermediate value of 5.8 for CA speakers. A repeated measures ANOVA on the mean scores assigned by the listeners to the three groups revealed a significant effect of speaker group,  $F(2,44) = 325.8$ ,  $p < 0.01$ , and Tukey *post hoc* tests indicated that all three groups differed significantly from each other at the 0.05 level. In

<sup>3</sup>In fact, including them in the analyses would have made no difference to the basic findings of the study.

summary, the CC group was rated as least Alabama-sounding, and the AA group was rated as most Alabama-sounding. The CA group received an intermediate rating that differed significantly from the ratings assigned to both the other groups.

As in Experiment 1, the listeners all showed a clear tendency to distinguish among the three groups of talkers. Every listener showed the same ordering of mean ratings for the three talker groups, and individual ANOVAs on each listener's data revealed that all but 5 of the listeners reached our stringent criterion of  $p < 0.002$ . The remaining listeners showed levels of  $p = 0.01$  or less.

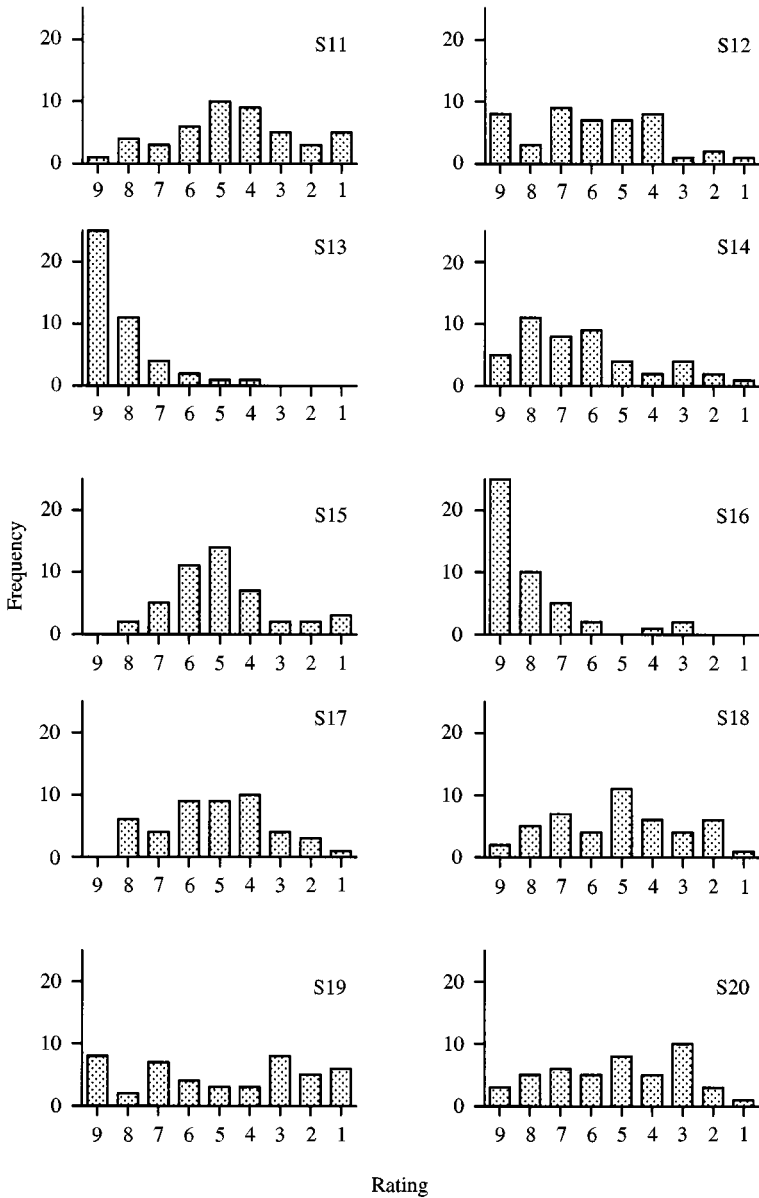
A one-way ANOVA on the data from the individual talkers, calculated for each talker by averaging over the ratings obtained for all the listeners, once again revealed a significant effect of talker group,  $F(2, 27) = 31.1$ ,  $p < 0.01$ . *Post hoc* Tukey tests indicated that the mean score assigned to every group differed significantly from the mean for every other group ( $p < 0.05$ ).

Fig. 4 shows the individual histograms of the CA speaker group for the Alabama listeners. The ratings of speakers 13 and 16 were remarkably comparable to the ratings assigned by the Canadian listeners. Most ratings fell between 7 and 9, indicating that the listeners perceived these talkers to speak a variety of English that was considerably different from the Alabama norm. Talker 15, who received many ratings in the 4–6 range from the native Canadians, received a very similar distribution of scores from the Alabama listeners. Two CA talkers, 18 and 19, had received the largest number of ratings in the 6–9 range (American-sounding) from the Canadian listeners; they were also among the talkers who were most often assigned ratings near the Alabama end of the continuum by the Alabama listeners. The Canadian immigrant speaker (19) who was judged to be the most American-sounding by the Canadian listeners ( $M = 6.7$ ) was also ranked highly (second on the scale) by the Alabama listeners. The Canadian immigrant speaker (11) who received the lowest mean rating from the Alabama listeners (indicating that he sounded most like an Alabaman speaker) had a mean rating lower than that of 2 of the AA speakers. The Canadian listeners, on the other hand, ranked 6 other CA speakers as more American-sounding than this individual. Despite this apparent disagreement, however, the pattern of the histograms suggested that the ratings assigned by the two listener groups are more alike than they are different.

One additional analysis was carried out to determine whether the two groups of listeners tended to rate the CA group as closer to the Alabaman or Canadian end of the rating scale. Single-sample *t*-tests (two-tailed) on the listener data revealed that the mean ratings assigned by both groups differed significantly from 5 ( $M = 3.99$  for the Canadian listeners;  $M = 5.84$  for the Alabaman listeners). In other words, despite slight differences in the wording of the scales, both groups of listeners rated the CA talkers as more Canadian than Alabaman.

#### 4. Experiment III

Experiments I and II established that the native listeners from both Canada and Alabama were able to perceive phonetic differences among the three groups of speakers. Our next step was to evaluate the speech samples to identify some of the stimulus properties that may have resulted in the observed differences in ratings. Our goal was not to carry out a detailed phonetic analysis of the samples, but merely to verify that some specific phonetic properties of the CA speakers' productions were indicative of dialect



**Figure 4.** Distributions of the Alabaman listeners' ratings for the individual speakers in the Canadian immigrant group. The horizontal axes have been reversed to permit easy comparison with Fig. 2.

learning. We began with our own impressionistic analysis, taking into consideration the comments elicited from our listeners in debriefing, and then selected two phonetic variables to examine in greater detail. To ensure that our own biases (due to our familiarity with the speech samples) did not lead us to misanalyze the speakers' productions, we carried out a new set of assessments with two phonetically trained listeners who were blinded as to the group membership of the 30 talkers examined in the study.

Some of our listeners thought that the Alabama speakers spoke more slowly than the Canadians. This may merely reflect a stereotyped notion about southern U.S. speech. On the other hand, the listeners may have used articulation rates as a basis for their judgments. We therefore analyzed the articulation rates in the speech samples.

Our own evaluations of the stimuli led us to believe that the most salient segmental difference between the CC and AA speakers had to do with their productions of diphthongs. For example, in the words “wife”, “like”, and “rifle”, the CC speakers always produced the “raised” Canadian variant [ʌj]. The AA speakers, of course, never did. In some instances they produced a diphthongal [aj], but in several of their productions of these words and in “highway”, “driving”, and “goodbye”, they produced a more monophthongal [a]. Because of the content of the cartoon we used for speech elicitation, one or more of these six words occurred in the narratives of most of the speakers. It was therefore feasible to excise tokens of these words for further evaluation by two trained listeners.

Another reason for focusing on instances of [aj] in this experiment was that there was little, if anything, else at the segmental level that struck us as a reliable indicator of the speakers’ backgrounds. However, at the prosodic level we noted that many of the AA speakers tended to use a distinctive phrase-final intonation pattern involving lengthening and a flat, relatively high pitch. It was beyond the scope of this study to attempt to quantify this property of the speakers’ productions or to have our expert listeners evaluate prosodic features independently of segmental properties.

#### 4.1. Method

##### 4.1.1. Procedure

To determine whether the Canadians spoke more quickly than the Americans, we calculated articulation rates for each speaker from the CC and AA groups by measuring total speaking times, not including pauses, false starts, repetitions, and discourse particles. All measurements were made to the nearest 1 ms with a waveform editing program. Articulation rates (syll/s) were obtained by dividing the total number of syllables uttered by the total speaking times.

For the vowel analysis we attempted to find at least one suitable token of one of the six words “wife”, “rifle”, “like”, “driving”, “highway”, and “goodbye” from as many speakers as possible. However, not every speaker produced one of the target words, and because the speech samples were extemporaneous, some tokens were very reduced and might not have been intelligible if removed from context. In the end, we found useable tokens from nine speakers in the AA group, six speakers in the CC group, and nine speakers in the CA group. However, we were able to find two different words (of the six under consideration) from some of the speakers to create a stimulus set of 10 words from each group. We excised the relevant words using a waveform editing program and saved them as 16-bit audio files.

For the evaluations, we recruited two phonetically trained Canadian listeners with graduate degrees in linguistics and experience in teaching undergraduate phonetics classes. The listeners assigned accent ratings independently during separate listening sessions in which they were blinded as to each speaker’s place of origin. They rated each word on a 5-point scale ranging from Very American (1) to Very Canadian (5). They heard the stimulus items through headphones in random order while following on

a printed page that listed each target word along with a labelled rating scale. We instructed them to listen for Canadian Raising (as is typically found in the Canadian pronunciation of 'wife', 'rifle' and 'like') and to rate those words with clearly raised vowels as 'Very Canadian'. They were to rate words with monophthongal /a/ (as is typically found in the Alabaman pronunciation of 'driving', 'highway' and 'goodbye') and words lacking raising where it would be expected in Canadian speech as 'Very American'. Other points on the scale were to be used for varying degrees of Canadian or American accent. The listeners heard the entire list of words once before making their judgments and were allowed to listen to each item as many times as they desired. After rating the full set, they heard the entire list again and were invited to make further changes to their ratings. The presentations were controlled by the experimenter (MJM), who presented them by number and so was unaware of which specific token the listeners were hearing at any time. In the case of one token of "goodbye" from the AA group, both listeners commented that the token sounded clipped and that they had difficulty evaluating it. We therefore decided to drop this token from the analysis.

#### 4.2. Results

The mean articulation rates for the CC and AA groups were 4.56 and 4.49 syll/s, respectively. This difference did not approach statistical significance,  $t(18) = 0.239$ ,  $p = 0.81$ .<sup>4</sup>

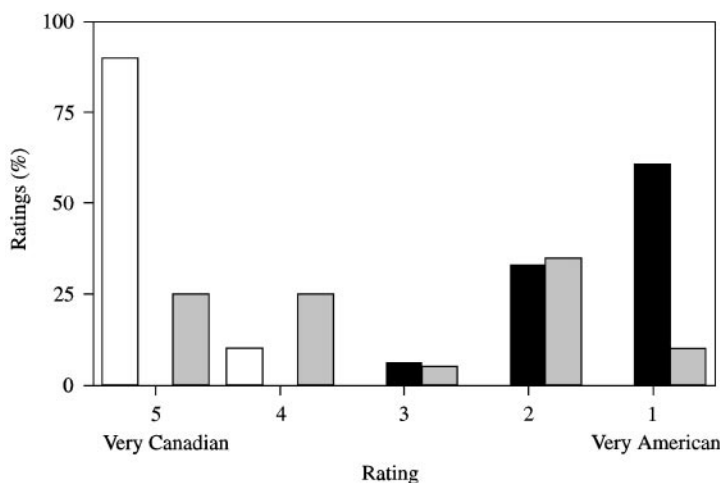
Fig. 5 shows the distribution of the ratings of the individual words for both listeners. While the ratings of the CC and AA tokens are clustered at the ends of the scale, the ratings of the CA tokens are spread across it. Half the ratings assigned to the CA tokens were scores of 1 or 2 (i.e., toward the "American" end of the scale).

#### 4.3. Discussion

We found no evidence of a systematic difference in speaking rates between the AA and CC groups. However, as noted earlier, we observed that many of the AA speakers used a distinctive phrase-final intonation pattern. This may have given the listeners the impression that the AA speakers articulated more slowly overall. While none of the CC speakers used this pattern, at least 2 of the speakers in the CA group appeared to do so to some degree. This tentative finding of the acquisition of prosodic features of D2 by our adult migrant group is an interesting issue worthy of exploration in future research.

The pattern of listener ratings shown in Fig. 5 is remarkably similar to the one seen in Figs 1 and 3. It seems clear from these data that even the isolated single-word tokens provided the raters with enough information to detect some degree of dialectal change in the CA group. The results lend support to our impression that changes in the production of the diphthong /aj/ probably contributed to the listeners ratings in Experiments I and II. In fact, given the degree to which the expert listeners were able to distinguish the speaker groups by this single segmental property, it is conceivable that this was the primary feature used by the untrained listeners in their ratings.

<sup>4</sup>We performed a parallel analysis on speech rates (including pauses). Again there was no significant difference between the two groups.



**Figure 5.** Distributions of ratings from two phonetically trained listeners for productions of the words “like”, “wife”, “rifle”, “driving”, “highway”, and “goodbye”. □, CC; ■, AA; ▒, CA.

## 5. General discussion

In this study we explored perceptible D2 speech learning in the productions of adult Canadians living in Alabama. A number of logically possible outcomes might have been observed in the three experiments reported here. One possibility was that the migrant Canadians living in Alabama would exhibit no characteristics of Alabama speech in their productions, in which case there would be no evidence of dialect change for our listeners to hear, and we would have observed negative results in all three experiments. Another possibility was that the migrants’ productions would exhibit characteristics of Alabama speech that would be detected by the trained listeners but that would not be noticeable to one or both of the groups of untrained listeners who participated. Neither of these outcomes was observed. Instead, in all three experiments we found evidence of perceptible D2 speech learning. In Experiments I and II, two groups of listeners, one from Canada and the other from Alabama, reliably rated the immigrants’ speech as having an intermediate degree of “American accent”, that differed significantly from that of the native Canadian and native Birmingham comparison groups. Given that judgments of the written forms of the narratives failed to yield any significant differences, it seems clear that the listeners’ evaluations were based on the phonetic properties of the speakers’ utterances and not on lexical choices or syntactic structures. A further analysis of specific properties of the migrant Canadians’ speech in Experiment III allowed us to conclude that at least some of them had adopted segmental, and possibly prosodic, characteristics typical of southern U.S. speech.

A wide range of factors probably influence the extent to which migrants acquire a D2. In this case, however, there were several reasons to expect that dialectal change in the migrant group would be minimal and perhaps not even noticeable to listeners. First, all the speakers were 18 years of age or more when they moved to Birmingham. If they had passed a sensitive period beyond which speech learning, especially of subtle phonetic phenomena, becomes difficult, they may not have been able to acquire Alabama



English. Second, the CA participants had little or no need to modify their speech since they already spoke a variety of English that was highly intelligible to Alabama English speakers. Third, the D2 in question may be regarded by many Canadians as a low-prestige variety of English; thus Canadians in Birmingham might actually be fairly resistant to adopting the D2. Finally, the speech samples were elicited by a native speaker of Canadian English. To the extent that the talkers accommodated their speech to that of the experimenter, they may not have manifested the full range of the southern American speech properties that they had actually acquired.

Our finding of perceptible D2 speech learning, in spite of these factors that might work against dialect change, provides new empirical support for one of the key postulates of Flege's (1995) Speech Learning Model: that the production and perception of speech sounds remain subject to adaptation across the life span. Although previous work has yielded evidence of speech learning in adult speakers of a second language, our data provide evidence for perceptible changes in first language speech production. Even adults who have passed the supposed critical period for language acquisition can modify their native language productions as a result of experience in a new dialect area. Theoretical perspectives that often emphasize what adults appear *not* to achieve in terms of L2 phonetic learning must account for this malleability even in L1 speech production. Of course, the extent to which speech modifications are under the conscious control of the speakers and the complex factors that may explain individual speaker differences remain to be explored.

Our findings are also of sociolinguistic importance. The fact that dialectal change in our migrant group was evident to linguistically untrained listeners representing both D1 and D2 indicates that such listeners may be in a position to make social judgments about migrants on the basis of dialectal change. Thus, migrants may be viewed more or less favorably by speakers of D1 or D2, depending on the extent to which they have modified their speech patterns. Previous work on D2 acquisition has tended to rely upon linguists' impressionistic analyses of D2 speech. An experimental demonstration that naïve listeners are sensitive to dialect change in adults serves as an important complement to this stream of research.

Although both listener groups rated the immigrant Canadians as more Canadian-sounding than American-sounding, their sensitivity to differences between the two groups of Canadian speakers led them to rate the Canadian immigrants to Alabama as having an intermediate degree of American accent. Since the Alabama listeners were unlikely to have had much exposure to Canadian English, it is of interest that the two sets of ratings were so similar. Both groups tended to regard some of the same speakers as showing greater or less acquisition of the D2, and both groups rated at least one of the CA speakers as indistinguishable from the native Alabamian speakers. These findings show that just as listeners are able to identify foreign-accented speech on the basis of very little information (Flege, 1984; Flege & Fletcher, 1992; Flege et al., 1995), they are also sensitive to partially acquired dialects of native speakers of English, whether the listeners themselves are speakers of the talkers' D1 or D2. Linguistically unsophisticated judges can rate foreign accentedness reliably on a continuum; so too can they easily recognize different degrees of second dialect acquisition, even when they are presented with short samples (in this case, only 10 s) of speech. Our positive findings suggest that the speech-rating task used in this study may be a useful tool in further research on dialect acquisition by adult migrants.

Many of the speakers in the CA group had been in Alabama for only a few years. Given that L2 research has shown that length of residence in an L2 environment has a relatively small but significant effect on L2 speech learning, we might predict that after an extended stay of many years, their speech will come to resemble more closely the D2. However, if D2 acquisition patterns do indeed parallel those of L2, we would expect that native speakers of Alabama English would be able to detect subtle differences (from the D2) in the speech of adult Canadian immigrants no matter how long they have lived in the D2 region. Just as adult second language learners tend to exhibit perceptible L1 accents in their L2, it is possible that most adult D2 learners are unlikely to ever sound like native speakers of D2.

The finding in Experiment III that our trained listeners could distinguish the three speaker groups on the basis of a single segmental characteristic (i.e., the diphthong /aj/) raises the important question of what it means to say that someone has acquired (from a phonetic standpoint) a second dialect. Although the differences between Canadian and Alabaman English were easily recognized by untrained listeners, the actual number of salient segmental and prosodic differences between these two varieties appears to be quite small, and it is possible that our untrained listeners attended to only one or two properties of the speech samples when making their judgments. This may explain why some of the migrants were judged to be just as "American-sounding" as the Alabama natives by both groups of listeners. It would be premature to conclude, however, that the acquisition of one or two characteristics of a D2 is generally enough to convince listeners of second dialect acquisition. The speech samples used in this study were very short and obviously did not convey information to the listeners about the full range of segmental and prosodic phenomena that might distinguish the two dialects in question. Further work with longer samples and a larger number of speakers is clearly needed.

Another promising area for investigation of the degree of D2 acquisition is the attitude of the migrant group towards the host culture. Unlike L2 learners, immigrants to a new dialect region may have little need to acquire D2 characteristics, except where there are problems with intelligibility across the dialects in question. In future work, an examination of the relationships among attitudes, critical features for intelligibility, and listeners' ratings of D2 accent may yield interesting findings.

A report on the Canadian listener data was presented at the 1995 conference of the Canadian Acoustical Association. The authors wish to thank Susan Russell and Dwight Gardiner for their assistance with Experiment 3, as well as Alice Faber, two anonymous reviewers, and Pam Beddor, for many useful comments on the manuscript. We gratefully acknowledge the financial support of the Social Sciences and Humanities Research Council of Canada for a grant to the first author, and the University of Alberta for a Support of Academic Scholarship grant to the second author.

## References

- Anderson-Hsieh, J., Johnson, R. & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure, *Language Learning*, **42**, 529–555.
- Asher, J. & Garcia, R. (1969). The optimal age to learn a foreign language, *The Modern Language Journal*, **53**, 334–341.
- Brennan, E. & Brennan, J. (1981). Measurements of accent and attitude toward Mexican-American speech, *Journal of Psycholinguistic Research*, **10**, 487–501.
- Chambers, J. (1973). Canadian raising, *Canadian Journal of Linguistics*, **18**, 113–35.
- Chambers, J. (1991). Canada. In *English around the world: sociolinguistic perspectives* (J. Cheshire, editor), pp. 89–107. Cambridge: Cambridge University Press.

- Chambers, J. (1992). Dialect acquisition. *Language*, **68**, 673–705.
- Derwing, T. & Munro, M. (1997). Accent, intelligibility and comprehensibility: evidence from four L1s, *Studies in Second Language Acquisition*, **19**, 1–16.
- Flege, J. (1984). The detection of French accent by American listeners, *Journal of the Acoustical Society of America*, **76**, 692–707.
- Flege, J. (1995). Second language speech learning: theory, findings and problems. In *Speech perception and linguistic experience: issues in cross-language research*. (W. Strange editor), pp. 233–277. Timonium, MD: York Press.
- Flege, J. (1997). The role of category formation in second-language speech learning. In *New sounds 97. Proceedings of the 3rd international symposium on the acquisition of second-language speech* (J. Leather & A. James, editors), pp. 79–89. Klagenfurt, Austria: University of Klagenfurt.
- Flege, J. & Eefting, W. (1987). Cross-language switching in stop consonant production and perception by Dutch speakers of English, *Speech Communication*, **6**, 185–202.
- Flege, J. & Fletcher, K. (1992). Talker and listener effects on degree of perceived foreign accent, *Journal of the Acoustical Society of America*, **91**, 370–389.
- Flege, J., Frieda, E. & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2, *Journal of Phonetics*, **25**, 169–186.
- Flege, J. & Hammond, R. (1982). Mimicry of non-distinctive phonetic differences between language varieties, *Studies in Second Language Acquisition*, **5**, 1–18.
- Flege, J., Munro, M. & MacKay, I. (1995). Factors affecting strength of perceived foreign accent in a second language, *Journal of the Acoustical Society of America*, **97**, 3125–3134.
- Giles, H. & Smith, P. (1979). Accommodation theory: optimal levels of convergence. In *Language and social psychology* (H. Giles & R. St. Clair, editors), pp. 45–65. Oxford: Basil Blackwell.
- Hatch, E. & Lazaraton, A. (1991). *The research manual: design and statistics for applied linguistics*. New York: Newbury House.
- Krashen, S. & Seliger, H. (1975). Maturation constraints on second dialect acquisition, *Language Sciences*, **38**, 28–29.
- Labov, W. & Ash, S. (1997). Understanding Birmingham. In *Language variety in the South revisited* (C. Bernstein, T. Nunnally & R. Sabino Editors), pp. 508–573. Tuscaloosa, AL: The University of Alabama Press.
- Lippi-Green, R. (1997). *English with an accent*. New York: Routledge.
- Long, M. (1990). Maturation constraints on language development. *Studies in Second Language Acquisition*, **12**, 251–85.
- Major, R. (1992). Losing English as a first language, *The Modern Language Journal*, **76**, 190–208.
- Moyer, A. (1999). Ultimate attainment in L2 phonology: the critical factors of age, motivation, and instruction, *Studies in Second Language Acquisition*, **21**, 81–108.
- Munro, M. & Derwing, T. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners, *Language Learning*, **45**, 73–97.
- Munro, M., Flege, J. & MacKay, I. (1996). The effects of age of second-language learning on the production of English vowels, *Applied Psycholinguistics*, **17**, 313–334.
- Oyama, S. (1976). The sensitive period for the acquisition of a nonnative phonological system, *Journal of Psycholinguistic Research*, **5**, 261–285.
- Patkowski, M. (1990). Age and accent in a second language: a reply to James Emil Flege, *Applied Linguistics*, **11**, 73–89.
- Payne, A. (1980). Factors controlling the acquisition of the Philadelphia dialect by out-of-state children. In *Locating language in time and space* (W. Labov, editor), pp. 143–178. New York: Academic Press.
- Pulvermüller, F. (1995). What neurobiology can buy language theory: a response to Eubank and Gregg, *Studies in Second Language Acquisition*, **17**, 73–77.
- Pulvermüller, F. & Schumann, J. (1994). Neurobiological mechanisms of language acquisition, *Language Learning*, **44**, 681–734.
- Ryan, E., Carranza, M. & Moffie, R. (1977). Reactions toward varying degrees of accentedness in the speech of Spanish–English bilinguals, *Language and Speech*, **20**, 267–273.
- Sancier, M. & Fowler, C. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English, *Journal of Phonetics*, **25**, 421–436.
- Scovel, T. (1989). *A time to speak: a psycholinguistic inquiry into the critical period for human speech*. Cambridge: Newbury House.
- Tahta, S., Wood, M. & Lowenthal, K. (1981). Foreign accents: factors relating to transfer of accent from the first language to the second language, *Language and Speech*, **24**, 265–272.
- Thompson, I. (1991). Foreign accents revisited: the English pronunciation of Russian immigrants. *Language Learning*, **41**, 177–204.