Non-natives' Production of Vowels in Conversational Speech

Thorsten Piske*, James E. Flege*, Ian R. A. MacKay**, and Diane Meador***

*Rehabilitation Sciences, Univ. of Alabama at Birmingham, VH503, Birmingham, AL 35294
**Dept. of Linguistics, Univ. of Ottawa, 70 Laurier Ave E., Ottawa, ON K1N 6N5
***Speech and Language Sciences Research Laboratory, POB 20743, Juneau, AK 99802

Summary: This study investigated English vowels produced in conversational speech by 72 native Italian (NI) speakers who differed in their age of arrival (AOA) to Canada and/or in amount of L1 use. AOA, but not amount of L1 use, was found to affect L2 vowel production accuracy.

INTRODUCTION

In a recent article, Flege et al. (2) examined the production of English vowels by 72 NI subjects who differed in terms of AOA and/or in terms of their self-reported percentage use of Italian. Vowel production was elicited in two ways. The subjects repeated a series of four aurally presented real words, all containing the same vowel. When they later heard the same series of words a second time, they took the vowel occurring in all words and inserted it into a non-word frame (/b_do/). One important finding was that subjects in the Early (AOA = 7 yr) and Mid (AOA = 14 yr) groups but not the Late (AOA = 19 yr) group were able to produce the vowels /i u ø/ accurately in real words. However, none of the NI groups accurately produced these vowels - none of which occurs contrastively in Italian - in non-words. The purpose of this study was to further examine the production of English vowels by the same sample of NI subjects. Six English vowels produced in conversational speech were examined. Most researchers would acknowledge that conversational speech should represent the most important criterion for success in acquiring L2 vowels. However, no previous research has rigorously examined L2 vowels in conversational speech. The reason for this gap is that it is difficult to analyze such speech samples under controlled conditions. A new technique with satisfactory experimental control was developed for this study to analyze conversational speech.

METHODS

Talkers: The 90 subjects, who were all living in Ottawa, Canada, had a mean age of 48 years (SD = 6). The 18 native English (NE) control subjects had all been born and raised in Canada, whereas the 72 NI subjects had all been born in Italy. They had lived in Canada for an average of 35 years at the time of testing. The NI subjects were assigned to four subgroups of 18 subjects each. Three NI groups differed primarily according to mean AOA (Early-7, Mid-14, Late-19 years, see above). The Early group differed from another early bilingual group, called
“EarlyLo”, in amount of L1 use (EarlyLo-8%, Early-32%).

**Speech Materials:** The subjects responded extemporaneously to eight questions about topics of interest. Their responses were recorded on a DAT tape recorder, then later digitized at 11.025 kHz and transcribed orthographically. The orthographic transcriptions of the speech samples were used to identify words containing six English vowels of interest: /i, ə, u, i, æ, o/. Up to 18 words (3 words x 6 vowels) were identified per vowel for each subject. All the words were content words consisting of less than four syllables. Most importantly, the vowels to be rated always occurred in stressed syllables. Next, stretches of speech containing the words with the target vowels were edited out. The aim was to retain the syllable with the target vowel and at least one syllable before and after. However, only whole words were retained. For example, if the target vowel /u/ was found in a word like “difficult”, and if this word was preceded by the word “very”, the whole phrase (i.e., “very difficult”) was retained.

**Listeners:** Six native speakers of Canadian English who had all received training in the use of phonetic symbols were asked to evaluate a total of 1343 such stimuli. The listeners, who were from Ottawa, were between 22 and 53 years of age (M = 34 years).

**Procedures:** The listeners were tested in a quiet room at the University of Ottawa. Before participating in the goodness rating experiment they had to demonstrate an ability to accurately transcribe ten Canadian English vowels (/i, ɪ, ə, u, o, æ, ə, ɔ, U, ʌ/) and three Italian vowels (/a, e, ɛ/). In the goodness rating experiment that followed, the vowels /i/, /ə/, /u/, /ɪ/, /æ/, /o/ were presented separately over headphones in six randomized blocks, the order of which was counterbalanced across listeners. The stimuli were presented both aurally and visually. A written version of the phrase (or word) containing the target vowel appeared on the screen as it was presented auditorily. The target vowel was always marked on the screen by an asterisk so that there was no doubt as to what vowel was to be judged.

The listeners were asked to evaluate each vowel by pushing one of 15 buttons. The first three buttons were labeled “good”, “slightly distorted” and “very distorted” instances of the target vowel. In addition, these buttons all showed the phonetic symbol for the target vowel. The remaining 12 buttons showed the phonetic symbols used for all other Canadian English and Italian vowel categories (see above). If the listeners heard the vowel that was to be judged as an instance of its intended category, they were asked to rate it by clicking one of the first three buttons. If they did not hear this vowel as an instance of its intended category, they were asked to transcribe it by clicking the button which represented the vowel they had heard.

**RESULTS**

The first question addressed here was whether the NI subjects’ vowels would diverge increasingly more from those of the NE subjects’ vowels as AOA increased. The mean ratings obtained for the subjects were submitted to a (4) Group x (6) Vowel ANOVA. As shown in Fig. 1, the ratings for all six vowels decreased as AOA increased, although at somewhat different rates. The mean ratings obtained for the NE subjects were compared to the three NI groups who differed according to AOA (Early, Mid, Late). The ANOVA yielded a significant effect of Group, F(3,68) = 54.0, p < .001, and two-way interaction, F(15,340) = 3.2, p < .001.
The simple effect of Group was significant for all six vowels \( (p < .001) \). Tukey’s tests were carried out to determine which pair-wise between-group differences were significant at the .05 level. The NE group’s productions of all six vowels received significantly higher ratings than did the Mid and Late groups’ productions, but did not differ significantly from the Early group’s productions. The interaction might be attributed to the fact that the Mid group’s productions of just two vowels \(/\text{æ}/, /\text{u}/\) received higher ratings than the Late group’s. Differences in the pattern of between-vowel differences may also have contributed to the interaction. There were no significant differences for the Early subjects. The NE group’s \(/\text{æ}/\) received lower ratings than their productions of the other five vowels \( (p < .05 \text{ by Tukey’s test}) \); and, for the Mid group, \(/\text{æ}/\) received lower ratings than \(/\text{i}/\). Finally, for the Late group, \(/\text{i}/, /\text{ɒ}/\), and \(/\text{æ}/\) received lower ratings than \(/\text{i}/\) and \(/\text{ɒ}/\), and \(/\text{u}/\) received lower ratings than \(/\text{æ}/\) \( (p < .05)\).

**FIGURE 1.** The mean goodness ratings for English vowels spoken by a native English (NE) group and three groups of Italian/English bilinguals who differed in age of arrival in Canada \( (E = \text{Early}, M = \text{Mid}, L = \text{Late}) \). 1 = very distorted, 2 = slightly distorted, 3 = good production.

**FIGURE 2.** The mean goodness ratings for English vowels spoken by a native English (NE) group and groups of early Italian/English bilinguals who differed in amount of L1 use \( (\text{E} = \text{Early}, E = \text{Early}) \). 1 = very distorted, 2 = slightly distorted, 3 = good production.

The second question addressed here was whether there was an effect of L1 use on English vowel production. The mean ratings obtained for the NE subjects were compared to ratings obtained for two groups of early bilinguals in a \( (3) \) Group \( \times (6) \) Vowel ANOVA. These groups were matched for AOA, but differed according to self-reported use of Italian. As shown in Fig. 2, vowels spoken by the NE subjects obtained higher ratings than did those of the EarlyLo and Early subjects. The effect of Group was significant, \( F(2,51) = 10.7, p < .001 \), but did not interact with Vowel, \( F(10,255) = 0.7, p > .10 \). A Tukey’s test revealed that the two bilingual groups did not differ from one another \( (p > .10) \), but that both differed from the NE group \( (p < .01) \).

The final question addressed by this study was whether early Italian/English bilinguals can produce the “new” vowels \(/\text{ɪ}/, /\text{ʊ}/\) accurately in conversational speech. A Tukey’s test revealed that the early bilinguals did not receive significantly lower ratings for these three vowels than the
subjects in the NE control group ($p < .05$). This result, then, resembles the one obtained in (2) for vowels produced in a real-word repetition task, but not in a non-word production task.

DISCUSSION

This is the first study we are aware of that quantitatively examined L2 vowels produced in conversational speech. It showed that none of the six vowels examined here was produced in a native-like manner by the Mid and Late Italian/English bilinguals. This finding differs from the results obtained in the earlier study (2) for vowels produced in a real-word repetition task. In that task, all vowels except /i/ were produced inaccurately by the Late group, whereas subjects in the Mid group produced only the vowels /o/ and /u/ inaccurately. This suggests that a word repetition task might not adequately reflect how well Mid and Late learners produce some L2 vowels in conversational speech.

Another finding was that early bilinguals who continued to use their LI frequently did not receive lower ratings than early bilinguals who used their LI infrequently. This finding differs from the results obtained in previous studies investigating overall degree of foreign accent in the productions of short sentences (1, 3). In those earlier studies, early Italian/English bilinguals in a “high LI use” group were found to have significantly stronger foreign accents in English than early learners in a “low LI use” group. The fact that no LI use effect was found in this study might mean that there is too little phonetic information in a single vowel for listeners to note differences in degree of foreign accent. Consequently, the LI use effects reported in previous studies may reside in levels higher than a single segment.

Finally, we found that early Italian/English bilinguals can learn to produce the “new” vowels /u o/ accurately in conversational speech. This finding, as well as the result obtained in (2) that early bilinguals are also able to produce these three vowels accurately in a real-word repetition task, suggests that the early bilinguals’ representations’ of these vowels do not differ from the native speakers’ representations. Thus, it remains to be determined why they were not able to produce /u o/ accurately in non-words. Further research is needed to resolve this question.

ACKNOWLEDGMENTS

This research was funded by a grant from the National Institute of Deafness and Communication Disorders (DC00257).

REFERENCES