# Investigating native and non-native vowels produced in conversational speech

*Thorsten Piske[1], James E. Flege[2], Ian R. A. MacKay[3] and Diane Meador[4]*
*[1] University of Education Schwäbisch Gmünd, Germany; [2] University of Alabama at Birmingham, USA; [3] University of Ottawa, Canada; [4] University of Alaska Southeast, USA*

ABSTRACT

The primary aim of this study was to determine if errors observed when fluent early bilinguals produced L2 vowels in non-words would also occur when the same bilinguals produced vowels in conversational speech. The subjects were native speakers of Italian who began learning English when they immigrated to Canada as children or adults ("early" vs. "late" bilinguals). The early bilinguals were subdivided into groups differing in amount of continued L1 use ("Early-low" vs. "Early-high"). Vowel production accuracy was assessed auditorily by native English-speaking listeners. The listeners knew beforehand the identity of the target vowels being evaluated. The experiment revealed that vowel errors produced in non-words by early bilinguals who continued to use their L1 Italian relatively often were not apparent in conversational speech. The findings reported here support the view that L2 production data elicited in experiments involving the use of written materials may not always reveal how accurately bilinguals can produce the sounds of an L2 in conversational speech.

KEYWORDS: L2 vowel production, conversational speech, spelling pronunciations

## 1. Introduction

The research reported here was motivated by the results of a study by Piske et al. (2002), who examined the production of 11 English vowels (i.e., /i ɪ eɪ ɛ æ u o ʌ ɒ ʊ ɚ/) by 54 native Italian subjects differing in terms of age of arrival to Canada (AOA) and/or in terms of their self-reported percentage use of Italian. Piske et al. (2002) elicited vowel production in two different ways: The subjects first repeated a series of four real English words that were modelled aurally and which also appeared on a written list. When the same four words were presented a second time, the subjects inserted the vowel common to all four (e.g, /i/ in *read*, *deed*, *heed*, *bead*) into a /b_do/ frame, creating a non-word (e.g., /bido/). 11 native speakers of English were recruited to auditorily evaluate the 1584 vowel stimuli that had been produced by three groups of Italian-English bilinguals and one group of native English (NE) speakers. Before the stimuli were presented to the

native English-speaking listeners they had been edited to minimize the possibility that lexical factors would influence the listeners' judgments. Piske et al. (2002) found that whereas none of the vowels produced by a group of early bilinguals who seldom used their L1 Italian (i.e., the "Early-low" group) received significantly lower ratings than vowels produced by the NE subjects, some of the vowels (i.e., /ɪ ɛ ɚ ʌ ʊ/) produced by a group of early bilinguals who used their L1 frequently (i.e., the "Early-high" group) did receive lower ratings than vowels spoken by the NE subjects. Interestingly, most of the observed differences between the NE and Early-high groups were for vowels spoken in the non-word condition. In a second experiment included in the Piske et al. (2002) study, native English-speaking listerners used keywords in order to classify the same vowels that had been rated in the first experiment. The results of the classification experiment suggested that many of the errors produced by the Early-high group were due, at least in part, to the influence of orthography. The vowels that were produced less accurately in non-words than in words by the Early-high group included the two vowels /ɪ/ and /ʊ/, which – as far as we know – do not have a phonetic counterpart in any variety or dialect of Italian. Table 1 shows how the NE and Early-high groups' productions of these two vowels in words and non-words were classified in the classification experiment reported by Piske et al. (2002).

Table 1. Native English listeners' classifications of 2 English vowels spoken by the subjects in two groups (adapted from Piske et al. 2002, p. 64).

| Group | Vowel | Words | Nonwords |
|---|---|---|---|
| NE | /ɪ/ | ɪ(95) ɛ(5) | ɪ(95) ɛ(4) |
| Early-high | /ɪ/ | ɪ(97) ɛ(3) | ɪ(35) i(57) |
| NE | /ʊ/ | ʊ(97) ʌ(3) | ʊ(74) u(15) ʌ(7) o(3) |
| Early-high | /ʊ/ | ʊ(94) ʌ(5) | o(34) ʊ(31) u(30) ɒ(3) |
| Only percentages greater than 2% are shown. | | | |

As summarized in Table 1, the /ɪ/s produced by the Early-high subjects in words were never classified as /i/, but the /ɪ/s they produced in non-words were classified as /i/ in 57% of instances. According to Piske et al. (2002), the bilinguals in the Early-high group appear to have accessed phonological representations for words stored in their mental lexicon when asked to produce the four real English words *rid*, *did*, *hid* and *bid*. However, when they were asked to insert the vowel common to these four words into a /b_do/ frame some of them appear to have pronounced the letter "i" in *rid*, *did*, *hid* and *bid* as it is pronounced in Italian (i.e., /i/). The errors involved in the production of /ʊ/ in non-words also appeared to be due, at

least in part, to the influence of orthography. All four words with /ʊ/ were spelled with at least one "o" (*good*, *could*, *would*, *hood*), which is often pronounced /o/ in Italian. The Early-high group's productions of /ʊ/ in words were never classified as /o/, whereas the /ʊ/s they produced in non-words were heard as /o/ in 34% of instances. Similar patterns suggesting an influence of orthography were also observed by Piske et al. (2002) for vowels other than /ɪ/ and /ʊ/.

The purpose of the experiment reported in this paper was to examine whether the spelling pronunciations observed for some vowels in the non-word condition would also occur in conversational speech samples. We hypothesized that errors observed in the Early-high subjects' productions of English vowels in non-words were an artifact of the elicitation procedure and that they did not typify how experienced Italian-English bilinguals produce English vowels when asked to answer questions in a relatively spontaneous way. If many of the Early-high subjects' errors for vowels in non-words were, in fact, an artifact of the elicitation procedure, two things should be true: First, the types of errors observed for the Early-high subjects' non-word productions of /ɪ/ and /ʊ/ should not occur in conversational speech samples. Second, their productions of other vowels in conversational speech should not differ from the NE subjects' productions of these vowels in conversational speech samples.

Most researchers would acknowledge that conversational speech *should* represent the most important criterion for success in acquiring L2 vowels, but surprisingly few studies have been undertaken (e.g., Wode 1981; Piske et al. 1999; Tsukada 2001; Bent et al. 2007). The most likely reason for this gap is the inherent difficulty in analyzing conversational speech under controlled conditions. A technique with satisfactory experimental control was used in this study to analyze conversational speech samples.

## 2. Method

### 2.1. Subjects

The same sample of subjects as those who participated in the Piske et al. (2002) study was used in the study presented here (see Table 2). The mean age of the 72 subjects who participated in the study was 48 years. Three groups of Italian-English bilinguals and a group of NE speakers (18 per group) were recruited in Ottawa, Ontario.[1] The native Italian subjects had been living in Canada for a minimum of 18 years at the time of testing (mean = 35 years). The subjects in a group referred to as the "Late-high" group had arrived in Canada later in life

[1] Vowels spoken by 18 native speakers of Italian with a mean AOA of 14 years were also elicited and evaluated by the listeners. In the interest of economy, the results of this "Mid" group will not be reported here.

(mean = 19 years) than two groups of early bilinguals had (mean = 7 years for both). The early bilinguals in a group referred to as the "Early-low" group reported using Italian much less often (mean = 8%) than the early bilinguals in an "Early-high" group did (mean = 32%). An ANOVA examining the native Italian subjects' self-estimates of percentage L1 use was significant $F(2,51) = 18.4$, $p < 0.01$. A Tukey's test showed that, as intended by the design, the Early-high and Late-high groups used Italian more frequently than the Early-low group ($p < 0.01$), whereas the Early-high and Late-high groups did not differ significantly ($p > 0.05$).

Table 2. Mean characteristics of the 18 subjects in each of four groups (adapted from Piske et al. 2002, p. 54).

| | Gender | Age | AOA | % Italian | LOR |
|---|---|---|---|---|---|
| Native English | 9m, 9f | 48 (7) | -- | -- | -- |
| Early-low | 9m, 9f | 48 (5) | 7 (3) | 8 (6) | 40 (5) |
| Early-high | 8m, 10f | 47 (6) | 7 (2) | 32 (16) | 40 (6) |
| Late-high | 8m, 10f | 48 (6) | 19 (1) | 41 (23) | 28 (5) |

Age = chronological age, in years; AOA = age of arrival, in years; % Italian = subjects' self-estimated percentage use of Italian; LOR = subjects's length of residence in Canada, in years. SDs are in parentheses.

## 2.2. Speech Samples

The following procedure was used to elicit conversational speech samples. The subjects heard a dialog between a husband and wife speaking English and Italian, respectively. The dialog was followed by alternating questions in English and Italian (four each). The NE and Italian subjects responded in English to the questions posed in English. The native Italian subjects also responded, in Italian, to the questions posed in Italian. The questions and the preceding dialog focused on issues like bilingualism and immigration. The questions were not designed to elicit the specific vowels of interest.

The subjects' responses to the English questions were recorded using a DAT tape recorder (Sony TCD-D8), and later digitized at 11.025 kHz. The digitized recordings were orthographically transcribed by an author who is a native speaker of English (i.e., the fourth author of the present paper). The transcriptions were then used to identify tokens of /ɪ/ /ʊ/ as well as of four additional vowels (i.e., /ɚ i æ o/) produced by the subjects in words consisting of 1 to 3 syllables.

An attempt was made to only select tokens occurring in content words. However, due to an insufficient number of content words with /ʊ/, tokens of /ʊ/ occurring in the modals *should*, *could* and *would* were also selected. All of the vowels included for analysis occurred in lexically stressed syllables.

Phonetic segments that are heard as intended in their original context may be misidentified when excerpted (e.g., Picket and Pollack 1963). We therefore preserved a short stretch of speech before and after each selected vowel token before storing it in a separate file. The preceding and following contexts both contained at least one syllable unless the syllable containing a target vowel was immediately preceded or followed by a pause. We avoided excerpting just parts of words. The contexts therefore included more than one syllable when this was necessary to preserve whole words. For example, in the excerpt "very d*i*fficult", the target vowel /ɪ/ was preceded by two syllables (the word *very*) and followed by two syllables (which completed the word *difficult*). An attempt was made to find three tokens per subject of all six vowels selected for analysis (/ɪ ʊ ɚ i æ o/). The average number of tokens available per vowel averaged 2.8 for the NE subjects, 2.4 for the Early-low subjects, 2.5 for the Early-high subjects, and 2.3 for the Late-high subjects. However, in some cases, no tokens of a particular vowel could be identified for certain subjects. This happened for 1 subject for the vowel /i/, 2 for /ɚ/, 4 for /o/, 7 for /æ/, and 18 for /ʊ/.

## 2.3. Auditory evaluation

Six native speakers of Canadian English (2 males, 4 females living in Ottawa, Ontario), having a mean age of 34 years, were asked to phonetically transcribe the target vowels occurring in the short excerpts that had been edited out of the conversational speech samples. Although all of the listeners had previously received training in the use of phonetic symbols, they were nevertheless required to demonstrate an ability to reliably transcribe the Canadian vowels /i ɪ ɛ æ u o ʌ ɒ ʊ ɚ/ and the Italian vowels /a e ɔ/ prior to participating in the experiment reported here.

During the experiment, tokens of each target vowel were randomly presented over headphones (Koss TD 65) in separate counterbalanced blocks (2 in each of 3 sessions). Blocking on vowel was intended to ensure that the listeners knew the intended identity of the target vowels (which were often, but not always obvious from the surrounding context). The listeners were told that they would hear short excerpts of English speech that had been produced by native speakers of English and Italian. They were told the identity of the target vowel to be evaluated in each block before it began. Each block began with five trials that were included for practice and were not analyzed.

An orthographic representation of each excerpt appeared on the screen as it was presented auditorily. The location of each target vowel was marked by an

asterisk in the orthography (e.g., "very d*fficult"). The listeners were told to focus their attention on the target vowel, and to identify it using one of the 15 response alternatives that appeared on the computer screen. For example, the responses offered for the target vowel /ɪ/ were: /i/, /e/, /ɛ/, /æ/, /a/, /u/, /ʊ/, /o/, /ɔ/, /ɒ/, /ʌ/, /ɚ/, "good /ɪ/", "slightly distorted /ɪ/", and "very distorted /ɪ/". This array of response alternatives permitted the listeners to identify the target vowel as being an instance of some other category or, when a vowel token was heard as intended, to rate it for goodness.

## 3. Results

In order to determine whether the types of errors observed when subjects in the Early-high group produced vowels such as /ɪ/ and /ʊ/ in non-words would also occur when the Italian-English bilinguals produced conversational speech, the percentage of times that each subject's production of the target vowels was heard as intended was calculated.[2] The means for each subject's productions of /ɪ ʊ ɚ i æ o/ were based on a maximum of 18 judgments (6 listeners x 3 tokens). The mean scores obtained for the four groups are shown in Table 3. The vowels spoken by subjects in the NE, Early-low, and Early-high groups were identified correctly more often (98.4-99.1%) than were the vowels spoken by the Late-high group (mean = 84.6%).

The standard deviations associated with the Late-high groups' scores were larger than those for the other three groups, so the scores were submitted to an arcsine transformation (Kirk 1968) before being examined in a series of one-way ANOVAs. As summarized in Table 3, the effect of group was significant for all six target vowels ($p < 0.05$). A series of five *a posteriori* t-tests were carried out to test for between-group differences for each vowel (NE vs. the three native Italian groups; Early-low vs. Early-high; Early-high vs. Late-high). Significantly lower scores were obtained for all six vowels produced by the Late-high than NE group (Bonferroni $p < 0.05$). No other between-group difference reached significance, however. As regards the intelligibility of the vowels examined here, these results confirmed the prediction that, for vowels produced in conversational speech, the Early-high subjects' productions of neither /ɪ ʊ/ /ɚ i æ o/ would differ from the NE subjects' productions of these vowels.

[2] This means that only the identification data but not the goodness ratings obtained from the native Canadian listeners were used in the analyses presented here.

Table 3. The mean percentage of times that English vowels spoken in conversational speech by subjects in four groups were heard as intended, and ANOVAs testing for between-group differences.

| | 1 Native English | 2 Early-low | 3 Early-high | 4 Late-high | *ANOVA* | *a posteriori* t-tests |
|---|---|---|---|---|---|---|
| /ɪ/ | 100.0 (0.0) | 99.7 (1.3) | 99.7 (1.3) | 77.3 (26.0) | F(3,68)=22.0* | 4 < 1, 2, 3 |
| /ʊ/ | 100.0 (0.0) | 99.1 (2.2) | 100.0 (0.0) | 90.1 (17.3) | F(3,50)=5.2* | 4 < 1, 2, 3 |
| /ɚ/ | 99.4 (1.8) | 99.7 (1.4) | 99.4 (1.8) | 85.8 (14.2) | F(3,65)=18.5* | 4 < 1, 2, 3 |
| /i/ | 99.1 (2.9) | 99.0 (2.2) | 99.1 (2.1) | 94.8 (5.5) | F(3,67)=6.1* | 4 < 1, 2, 3 |
| /æ/ | 96.6 (4.3) | 94.4 (8.1) | 94.8 (6.9) | 70.8 (18.1) | F(3,61)=16.8* | 4 < 1, 2, 3 |
| /o/ | 99.4 (1.8) | 98.6 (4.3) | 99.0 (4.0) | 89.7 (12.6) | F(3,64)=10.0* | 4 < 1, 2, 3 |
| *M* | 99.1 (1.0) | 98.4 (2.1) | 98.7 (1.2) | [illegible] (9.2) | | |

An asterisk indicates significance at the 0.05 level. The Bonferroni-corrected level used for the *a posteriori* tests was 0.05.

Table 4 summarizes the phonetic transcriptions obtained for the vowels spoken in conversational speech. The percentages shown here were calculated by dividing the number of times each phonetic symbol was used to identify a group's production of a particular vowel by the total number of available tokens. When the native Italian subjects' vowels were not heard as instances of the intended category, they were usually heard as the neighboring vowel (i.e., as a vowel adjacent in an F1-F2 space). This was expected from the results obtained by Piske et al. (2002) and in other studies (e.g., Flege et al. 1999). So, for example, intended /ʊ/s were sometimes heard as /u/, /ɚ/s as /ʊ/, /æ/s as /ɒ/ (or Italian /a/), and /o/s as Italian /ɔ/. The target vowel /ɪ/ was sometimes heard as /i/ and the target vowel /i/ was sometimes heard as /ɪ/.

Table 4. The mean percentage of classifications, by native English-speaking listeners, of six vowels that were spoken in conversational speech by the subjects in four groups.

| Group | Target Vowel | Classified as | Target Vowel | Classified as |
|---|---|---|---|---|
| Native English | /ɪ/ | ɪ(100) | /i/ | i(99.1) |
| Early-low | /ɪ/ | ɪ(99.7) | /i/ | i(99.0) |
| Early-high | /ɪ/ | ɪ(99.7) | /i/ | i(99.1) |
| Late-high | /ɪ/ | ɪ(79.2) i(19.5) | /i/ | i(94.2) ɪ(4.8) |
| Native English | /ʊ/ | ʊ(100) | /æ/ | æ(96.6) |
| Early-low | /ʊ/ | ʊ(98.6) | /æ/ | æ(95.0) ɛ(3.1) |
| Early-high | /ʊ/ | ʊ(100) | /æ/ | æ(94.4) a(3.0) ɛ(2.6) |
| Late-high | /ʊ/ | ʊ(88.9) u(11.1) | /æ/ | æ(71.4) ɒ(11.4) a(11.4) ɛ(4.3) |
| Native English | /ɚ/ | ɚ(99.4) | /o/ | o(99.4) |
| Early-low | /ɚ/ | ɚ(99.6) | /o/ | o(99.2) |
| Early-high | /ɚ/ | ɚ(99.4) | /o/ | o(99.6) |
| Late-high | /ɚ/ | ɚ(86.6) ʊ(6.3) ɒ(3.1) | /o/ | o(90.0) ɔ(7.0) ʌ(2.1) |

Percentages lower than 2.0% are not reported.

In their classification experiment, Piske et al. (2002) had found that the Early-high subjects' productions of /ɪ/ in non-words were classified as /i/ in 57% of instances (see Table 1). The /ɪ/s produced in words by the same subjects were never classified as /i/, however. The /i/-for-/ɪ/ substitutions in non-words were interpreted as resulting from orthography. As shown in Table 4, neither group of early bilinguals produced /ɪ/s that were heard as /i/ in conversational speech. This confirms the hypothesis that the early bilinguals' /i/-for-/ɪ/ substitutions in non-words did not typify their production of /ɪ/. Similarly, the Early-high groups' productions of /ʊ/ in words were never classified as /o/ in the classification experiment conducted by Piske et al. (2002), whereas their /ʊ/s were classified as /o/ in non-words in 34% of instances. This was also interpreted to be a spelling pronunciation. In support of this, the Early-high groups' productions of /ʊ/ in conversational speech were never classified as /o/.

## 4. Discussion

Taken together, the results of the Piske et al. (2002) study and the results reported here support the view that in certain elicitation conditions fluent early bilinguals may not produce all L2 vowels in a manner that is indistinguishable from L2 monolinguals. However, the differences in L2 vowel production between early bilinguals and L2 monolinguals are relatively subtle because the errors produced by early bilinguals in a specific elicitation condition may not be apparent when they produce L2 vowels in real words or in conversational speech. In the Piske et al. (2002) study, none of the vowels produced by a group of early bilinguals who only seldom used their L1 Italian were found to differ from the vowels produced by a group of NE subjects. Some of the vowels produced by a group of early bilinguals who continued to use their L1 Italian frequently were, on the other hand, found to differ from the NE subjects' vowels. Most of the observed differences between the NE and the Early-high groups were for vowels spoken in a non-word condition. The results of the classification experiment conducted by Piske et al. (2002) suggested that some of these errors were due to the influence of orthography. The results of the study reported in this paper confirmed the prediction that the errors observed in the Early-high subjects' productions of English vowels in non-words were an artifact of the elicitation procedure and that they did not typify how experienced Italian-English bilinguals produce vowels in conversational speech, because the vowel errors that had been produced by the subjects in the Early-high group in non-words were not apparent in the conversational speech samples analyzed here.

The results of a post-hoc analysis also provided indirect evidence that errors observed in non-words may not have typified the native Italian subjects' production of English vowels. We computed correlations between ratings of the native Italian subjects' overall degree of foreign accent that were obtained by Meador et al. (2000) and the percentage of times that six vowels produced by the native Italian subjects in words, non-words and conversational speech were identified correctly. Significant correlations were found to exist between degree of foreign accent in sentences and the percent correct scores for vowels spoken in all three contexts (words: $r(70) = 0.71$; conversational speech: $r(70) = 0.65$; non-words: $r(70) = 0.44$; $p < 0.01$). However, the correlations were significantly weaker for the vowels in non-words than for the vowels in either conversational speech, $X(1) = 4.67$, $p < 0.05$, or words, $X(1) = 8.84$, $p < 0.01$.

Although the subjects in the Early-high group examined here and in the Piske et al. (2002) study had already been living in a predominantly English-speaking environment for many years when they were tested, some of their errors in the production of vowels such as /ɪ/ and /ʊ/ in non-words still reflected the influence of orthographic input. The observation that these errors did not occur when the same subjects produced conversational speech supports the view (see, e.g., Piske

et al. 2001) that it is generally problematic to include written materials in experiments designed to examine how accurately bilinguals can produce an L2 (recall that the subjects examined by Piske et al. (2002) produced non-words after repeating four real English words that were modeled aurally and which also appeared on a written list). Additional research is needed, however, to determine why the grapheme-phoneme correspondence rules of Italian appeared to influence the Early-high group's production of English vowels only in non-words, but not in words and under what specific circumstances the grapheme-phoneme correspondence rules of a bilingual's L1 continue to exert a long-term influence on her/his production of an L2. As Piske et al. (2002) point out, the insertion of vowels into a /b_do/ frame resulted in non-occurring words in English, but the resulting non-words did resemble some Italian words (e.g., Carlson et al. 1985). According to Piske et al. (2002), it is possible that due to their continued frequent use of the L1 Italian the Early-high subjects' Italian lexicon was more strongly activated than the Early-low subjects' Italian lexicon so that their productions of the non-words were influenced to a greater extent by pronunciations specified in an Italian lexicon than were the Early-low subjects' productions of the non-words. It is up to future research to determine more precisely to what extent the production of L2 vowels in different elicitation conditions is indeed influenced by lexical factors and to what extent other mechanisms may be responsible for errors in the production of L2 vowels as they have been described in the present paper and by Piske et al. (2002).

## 5. Acknowledgments

This research was supported by grant no. DC00257 from the National Institute for Deafness and Other Communication Disorders.

# REFERENCES

Bent, T., Hargus Ferguson, S., Kewley-Port, D. 2007. Listener influences on vowel intelligibility. *Journal of the Acoustical Society of America* 122(5). 3016(A).

Carlson, R., Elenius, K., Granström, B., Hunnicutt, S. 1985. Phonetic and orthographic properties of the basic vocabulary of five European languages. *Speech Transmission Lab., Quart. Progress Status Rep.* 1. 63–94.

Flege, J. E., MacKay, I. R. A., Meador, D. 1999. Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America* 106(5). 2973–2987.

Kirk, R. 1968. *Experimental design: Procedures for the behavioral sciences*. Belmont: Brooks/Cole Publishing.

Meador, D., Flege, J. E., MacKay, I. R. A. 2000. Factors affecting the recognition of words in a second language. *Bilingualism: Language and Cognition* 3. 55–67.

Picket, J. M., Pollack, I. 1963. Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt. *Language and Speech* 6. 151–164.

Piske, T., Flege, J. E., Mackay, I. R. A., Meador, D. 1999. Non-natives' production of vowels in conversational speech. *Collected papers from the 137th Meeting of the Acoustical Society of America and the 2nd Convention of the European Acoustics Association: Forum Acusticum.* Paper no. 2ASCB4.

Piske, T., MacKay, I. R. A., Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2. A review. *Journal of Phonetics* 29. 191–215.

Piske, T., Flege J. E., MacKay, I. R. A., Meador, D. 2002. The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica* 59. 49–71.

Tsukada, K. 2001. Native vs. non-native production of English vowels in spontaneous speech. An acoustic study. *EUROSPEECH 2001 Scandinavia, 7th European conference on speech communication and technology, 2nd INTERSPEECH event*, Aalborg. 305–308.

Wode, H. 1981. *Learning a second language: An integrated view of language acquisition*. Tübingen: Narr.