

# The Acquisition of Swedish Long vs. Short Vowel Contrasts by Native Speakers of English, Spanish and Estonian

Robert McAllister\*, James E. Flege† and Thorsten Piske‡

\**Department of Linguistics, Stockholm University, Stockholm, Sweden*

‡*Department of Rehabilitation Sciences, University of Alabama at Birmingham, USA*

## ABSTRACT

This work addresses an important current issue in the field of second language speech acquisition. The main purpose of the experiments reported here has been to test a hypothesis about the influence of L1 phonology on the acquisition of contrastive L2 phonetic categories. This hypothesis holds that an L2 contrastive category will be difficult to acquire if it is based on a phonetic feature not exploited in the L1. Twenty native speakers each of American English, Latin American Spanish and Estonian as well as a control group of 20 native speakers of Swedish were given a production and perception test to assess their mastery of the Swedish long-short vowel contrast. Our results support our hypothesis indicating that the success of learning the Swedish long-short vowel contrast seems to be related to the role of the duration feature in the L1.

## 1. INTRODUCTION

A primary issue in both past and current studies of second language (L2) speech acquisition is how and to what extent the first language (L1) influences the learning of L2. This paper is concerned with the acquisition of a long-short contrast in L2 phonology. A familiar theme in the study of this contrast and one that has been of considerable interest in L2 acquisition research is that of the phonetic realization of the vowel length distinction which can be found in many languages. The experiments reported here are designed to test what could be called a "feature hypothesis" relevant to the learning of L2 phonology in general and to the learning of an L2 long-short vowel contrast in particular. This hypothesis was first mentioned in Flege's Speech Learning Model (SLM), one of the current models of L2 speech acquisition [2]. The hypothesis could be stated as follows: L2 features not used to signal phonological contrast in L1 will be difficult to perceive for the L2 learner and this difficulty will be reflected in the learner's production of the contrast based on this feature.

There is increasing evidence that features used in the L2 but not in the L1 present a problem for the L2 learner and that the use of these features in the L2 is age dependent. Gottfried and Suiter [3] found that Americans had little trouble learning Mandarin vowel quality but were less successful in learning lexical tone, a feature not found in the L1.

It is well documented that languages differ considerably with regard to the use of duration and spectral features in the phonological distinction we are concerned with here. Traditionally, the distinction has been attributed to

durational differences in the vowels, hence the long-short terminology. Many languages, of course, do not use duration as a cue to distinguish phonetic categories and consequently, native users of those languages, such as Spanish, have no experience with the use of durational cues [7]. Among the languages that do signal contrast with duration there are language specific relationships between the durational and spectral features in the realization of that contrast. For example, experimental evidence shows that in standard French duration plays only a minor role in the phonological system [6]. In English, spectral information covaries with the temporal cues and many studies have shown that native speakers of English are sensitive to the durational feature and are, in some circumstances, able to identify a vowel on the basis of its length. In contrast to Spanish, where duration seems to be of little or no consequence for identification of phonological distinctions, Estonian could be cited as an example of a language in which a complex pattern of durational relationships is a salient feature of the phonology and where vowel quality is of little or no consequence to the contrasts signalled by temporal cues [5]

The target L2 in this study is standard Swedish which has a durational contrast in the vowel system. The phonetic realization of this contrast in Swedish makes this language rather convenient in a test of the above stated hypothesis. Swedish words containing high or low vowels in the stressed syllable use both duration and vowel quality to signal the phonological contrast [4]. In this context, spectral information seems to be the most important cue for native listeners. In words with mid vowels, however, the contrast seems to be cued primarily by temporal differences.

By investigating the acquisition of a durationally based phonological contrast in an L2 using several of the languages cited as examples, it is our intention to perform a preliminary test of the what we have called the feature hypothesis.

There appear to be other phonetic facts relevant to a discussion of vowel length perception. Bohn [1] cites phonetic research to support the suggestion that language independent acoustic/auditory factors may play an important role in the perception of L2 vowel length. Other results in L2 experimental research, including some results from the studies of Koreans learning English mentioned earlier [3], may be interpreted as support for this suggestion. However, the results of a number of studies, when taken in combination, suggest that another explanation based on L1-L2 relationships may be more plausible. As mentioned earlier. The SLM [2] hypothesizes that the use of features not exploited in the L1 for the production and perception of L2 contrasts becomes less likely as the age of learning the L2 increases.

The feature hypothesis stated in this paper is clearly based on the idea that L2 speech learning is highly language specific at the phonetic level. According to our feature hypothesis, learning vowel length in Swedish for the high and low vowels should not be difficult for most L2 learners since all languages use spectral features in their differentiation of phonetic vowel categories. The length contrast should, however, for the mid vowels, be easier for those whose L1 phonology exploits durational cues.

## 2. METHODS

### 2.1. Subjects

A total of 80 subjects participated, first in a production experiment, and then in a perception experiment. All of the subjects were tested in a sound attenuated room at the Department of Linguistics at Stockholm University, and were long time residents of Sweden. 20 of these subjects were native speakers of Swedish. There were also 20 native speakers each of South American Spanish, North American English and Estonian. Roughly half of the subjects in each group were female. All of the Spanish speaking, English speaking and Estonian speaking subjects had learned the L2, Swedish, while living in Sweden. To be included in one of these L2 learner groups subjects had to have arrived in Sweden after their 18<sup>th</sup> birthday, lived in Sweden for at least ten years at the time of testing and report using Swedish frequently.

### 2.2 Production experiment

The subjects produced 40 frequent two syllable Swedish words, all with lexical accent II. The words were recorded using professional quality equipment (Panasonic Professional Digital Tape Deck 3700, Sennheiser red dot MKE2 microphone). Half of the words had a phonologically long vowel, and the remaining half had short vowels. There were ten words each in four groups representing high vowels (/u:/, /ʉ/), rounded mid (/ø:/ /ø/), and unrounded mid (/ɛ:/ /ɛ/), and low vowels (/a:/, /a/). There were 20 nouns and 20 verbs. No words were allowed that could be paired minimally with another word in Swedish by changing the phonological vowel length only. The decision was made to avoid the use of orthography when eliciting production of the 40 test words. Had this been done, the orthography would have indicated the phonological length of the vowels in the test words. A definition task was used instead. The first author wrote, and then recorded, definitions of the 40 words. For example, to elicit "spade" [spɑ:də] (eng. shovel, spade) the subjects heard the definition "Ett verktyg, som man gräver i trädgården med." (eng. A tool used for digging in the garden.). To ensure that the subjects recognized the target word from the definition, a picture of the word was shown and the first letter or letters were also given, which would not reveal the identity of the vowel. The target word was said at the end of a short carrier phrase. The non native subjects were able to say the target word in most instances and a record was kept

of whether the subject needed help recognizing the desired words.

The recorded productions were digitized and the temporal measurements were made using Cool Edit. Because of the difficulty in measuring some vowels when preceeded or followed by glides, only four vowels from each of the 4 vowel contrast categories mentioned above were included in the study.

### 2.3 Perception experiment

The purpose of the perception experiment was to find out whether the subjects could determine if the 40 common Swedish words they had produced in the production experiment contained a phonologically short or long vowel.

**2.3.1 Stimuli.** A phonetician, who is a native speaker of Swedish recorded productions of the 40 test words. He also produced a set of 40 non-words formed by changing these vowels in the test words from short to long or from long to short. Acoustic analyses of these non words revealed that the vowels were accurately produced. That is to say that they were the same as the vowels in the real words.

**2.3.2 Procedure.** The 80 naturally produced stimuli were randomly presented a single time immediately following a definition. Each real word and the non-word paired to it by changing the phonological length of the vowel were presented following the same definition. The real word and its corresponding non-word were always located on separate halves of the test. The subjects were asked to click a "correct" or "incorrect" button as a response.

A total of four percent scores was computed for each subject, one for each of the four vowel contrast groups. Each score was based on 20 judgements, 10 for the real words, 10 for the non-words (where the correct response was "incorrect" because the word does not exist in Swedish).

## 3. RESULTS

### 3.1 Production

The aim of the analysis examining vowel duration was to determine if the three groups of non-native subjects would produce temporal contrasts between the four Swedish long-short pairs that were equivalent to those spoken by the native Swedish subjects. The mean values obtained for the native Swedes are plotted in all three panels of figure 1 and compared to the three non-native groups. The leftmost for native Spanish speakers, the middle panel for native English and the rightmost for native Estonians. The native Swedes produced a much larger temporal difference between phonologically long and short vowels (mean=230 vs. 148 ms) than the native Spanish subjects (169 vs. 143 ms). The native English subjects produced a substantial difference between the short and long members of all four vowel pairs (mean=215 vs. 159ms). However, their temporal differences were somewhat smaller than those of the native Swedish subjects in every instance. Finally, the native Estonians produced large long versus short temporal contrasts for the four vowel pairs (mean=228 vs. 139 ms).

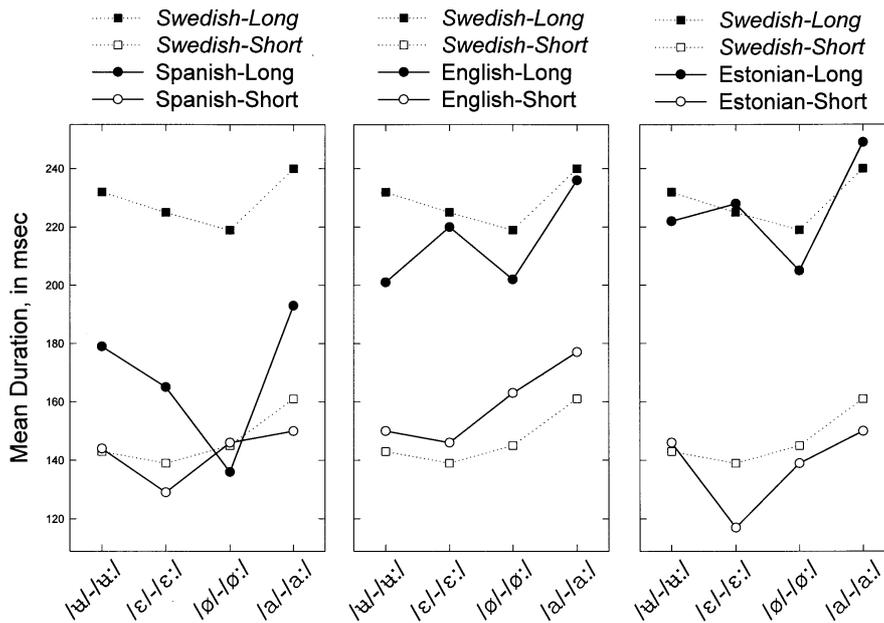


Figure 1 The mean duration of phonologically long and short vowels spoken by the native speakers of Spanish (left panel), English (middle panel), and Estonian (right panel), in msec. The mean values obtained for the native speakers of Swedish are juxtaposed to the values obtained for the non-native subjects in each panel

The mean duration values obtained for the 80 subjects were submitted to a three way ANOVA in which Group (4 levels) served as a between-subjects factor and Vowel Contrast (4 levels) and Phonological length (2 levels) as within subject variables ANOVA yielded a significant three-way interaction,  $F(9, 228)=2.99, p=0.002$ .

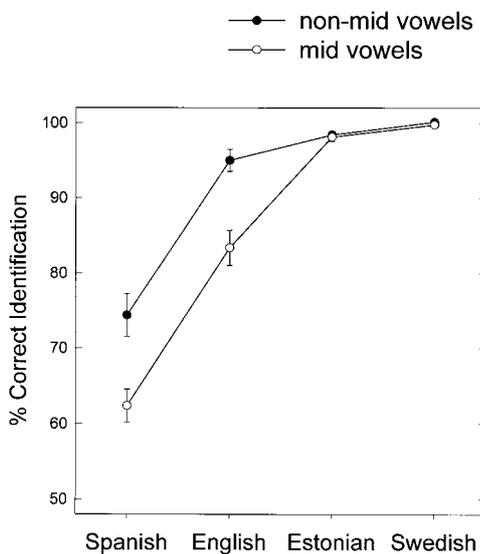


Figure 2 The percent correct scores obtained for two pairs of Swedish mid vowels and two pairs of Swedish non-mid vowels by the subjects in four groups.

### 3.2 Perception

Figure 2 shows the percentage of times that mid and non-mid vowels were recognized correctly. Like the Swedes, Estonian subjects obtained near-perfect scores. Scores for the native English subjects were lower on average (mid: 83%, non-mid: 95%), and those of the native Spanish subjects lower still (mid: 62%, non-mid: 74%). The difference between mid and non-mid contrasts for the native English and Spanish subjects, but not for the Estonian subjects, resulted in a significant interaction in an ANOVA examining the non-native subjects' scores,  $F(2,57)=13.8, p<0.01$ . For mid vowels, the Estonians' scores were significantly higher than the native Spanish subjects' scores ( $p<0.01$  by Tukey's test). For non-mid contrasts, the difference between the native Estonian and English subjects did not reach a significant level, with both groups obtaining higher scores than the native Spanish subjects ( $p<0.01$ ).

### 3.3 The production-perception relation

Figures 3 and 4 are scatter plots showing the relation between the production and perception scores for mid vowels (Figure 3) and non mid vowels (Figure 4). An X indicates the mean of the duration difference between the mid vowels and the non mid vowels. Note the large difference in the r-value: 0.70, for the mid vowels which was significant at the 0.0000 level and 0.41 for the non-mid, significant at 0.008 level.

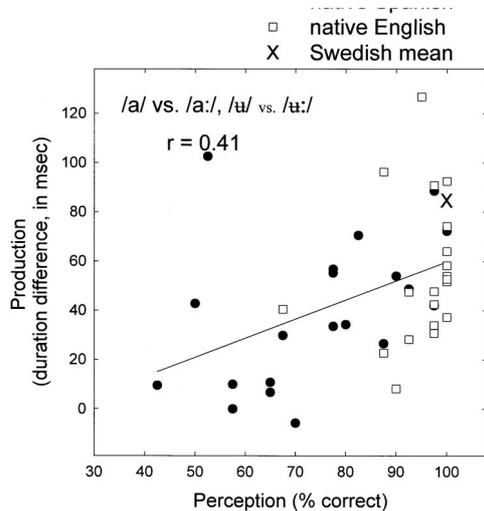


Figure 3 The relation between the size of temporal differences produced between the long and short members of two non-mid vowel pairs (/a/-/a:/, /ʉ/-/ʉ:/) and the scores obtained for these same vowels in the perception experiment. "X" indicates the mean duration difference observed in production and perception for the native Swedish subjects (mean production difference=80 msec, mean percent correct in perception=99.8).

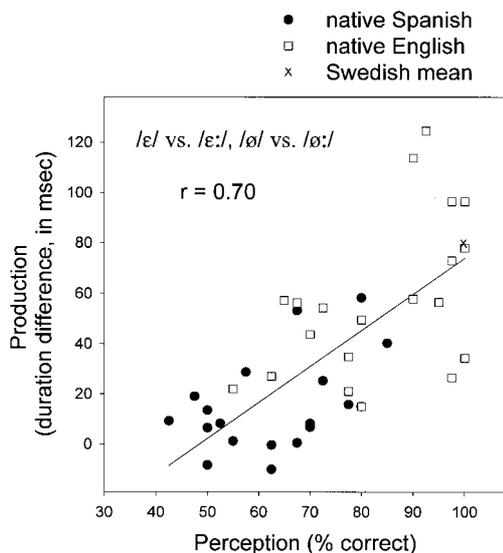


Figure 4 The relation between the size of temporal differences produced between the long and short members of two mid vowel pairs (/e/-/e:/, /ø/-/ø:/) and the scores obtained for these same vowels in the perception experiment. "X" indicates as in Figure 3 (mean production difference for Swedes=85 msec, mean percent correct in perception=99.9).

#### 4. DISCUSSION

A reasonable general interpretation of the results presented above would be that the native Estonian speaking subjects were most successful of the three subject groups in learning the long-short vowel contrast in Swedish. The native Spanish speakers appear to be least successful and the native English speakers somewhere in between but with very good success.

The feature hypothesis stated earlier predicts that the mid vowel long-short contrast based primarily on duration would be more difficult for the L2 learners whose phonology did not exploit this phonetic feature. This prediction seems to hold for the perception of these vowel categories in both native Spanish and English speakers. The Estonians, however, displayed high proficiency in both perception and production of the mid vowels. Miller and Grosjean [9], in their study of French dialects, postulate, on the basis of their experimental results, that the "overall prominence of a given acoustic property" in the L1, in this case duration, has important consequences for learning L2 phonology. While the notion of overall prominence could be considered vague, this interpretation could be applied to the results of the present study and used as support for the feature hypothesis. Further, Bohn's suggestion that those speakers who were not "sensitized" for the use of duration by their L1 (the native Spanish speakers in our study) used language independent acoustic/auditory processing strategies does not seem to be supported. The dependence of L2 speech learning on L1-L2 language specific phonetic relationships is seen not only in the poor performance of the Spanish speaking subjects but also in the relative success of the English speakers, whose L1 exploits duration. The overall prominence of duration in the prosody of Estonian would account for their success.

The differences in r-values in the correlation between perception and production scores could be interpreted as indicating that the Spanish and English speakers may have used spectral information instead of the durational information present for the learning of the non mid vowels. It appears, however, that the durational information may have been used in learning the contrast in the mid vowels.

#### REFERENCES

- [1] Bohn, S.-O. 1995. Cross language speech production in adults: First language transfer doesn't tell it all. In: Strange, W. red. *Speech perception and linguistic experience: Theoretical and methodological issues*. Timonium, MD: York Press.
- [2] Flege, J. (1995) Second language speech learning: Theory, findings and problems. In: Strange, W. red. *Speech perception and linguistic experience: Theoretical and methodological issues*. Timonium, MD: York Press.
- [3] Gottfried, T.L. and Suiter, T.L. 1997. Effect of linguistic experience on the identification of Mandarin tones. *Journal of Phonetics* 25: 207-231
- [4] Hadding-Koch, K. and Abramson, A. 1964. Duration Versus Spectrum in Swedish Vowels: Some Perceptual Experiments. *Studia Linguistica* XVIII II, 94-107.
- [5] Lehiste, I. 1997. The search for phonetic correlates in Estonian prosody. In Lehiste I. And Ross, J. (eds.) *Estonian Prosody: Papers from a Symposium*. Institute of Estonian Language and authors, Tallinn
- [6] Miller, J.L. and Grosjean, F. 1997. Dialect effects in vowel perception: The roll of temporal information in French. *Language and Speech*, 40, part 3, 277-288.
- [7] Navarro, T. 1968. *Studies in Spanish Phonology* Coral Gables, University of Miami Press