ABSTRACT
Cross-language research on adult speech perception demonstrates a strong effect of linguistic experience on consonant perception but not on vowel perception. Our paper re-examines the effect of linguistic experience on adults’ vowel perception. First, identification and goodness functions for the high front quadrant of the vowel space were mapped for speakers of Swedish, English, and Spanish. Second, speakers performed a discrimination task for one vector in this vowel space. Stimuli along this vector were identified by Swedish speakers as belonging to the Swedish front rounded vowel series /ç:/ - /ö:/ . However, English and Spanish speakers reported that the stimuli were not in their language. Significant differences in discriminability of these stimuli were observed across speakers of different languages. Our results show that linguistic experience plays a significant role in vowel discrimination.

1. INTRODUCTION
Research on speech perception shows that linguistic experience plays a substantial role in the perception of consonants in adults (Miyawaki et al., 1975; Iverson & Kuhl, 1996). The effect of linguistic experience on vowel perception, on the other hand, has seemed less robust in adults (although see Kuhl et al., 1992 for robust effects of linguistic experience on vowel perception in infants). In a study comparing English and Swedish speakers’ discrimination of front rounded vowels (phonemic in Swedish but not in English), Stevens et al. (1969) concluded that “[T]he listeners’ linguistic experience has essentially no effect upon their ability to discriminate small differences in vowel formant frequencies” (p. 12). However, task variables may have obscured the effects of language experience on vowel perception in that study.

Our goal is to better understand the relationship between linguistic experience, phonetic labeling, and acoustic sensitivity in vowels. We ask two main questions: 1) How do speakers of different languages map vowel space? 2) How is vowel discrimination affected by experience?

Figures 1, 2, and 3 display the front, high-mid corner of the vowel space. Mean production data from three different studies in three different languages: Swedish, English, and Spanish are reproduced here. Note that below F2 = 1900 Hz, Swedish has two vowels, but English and Spanish have none. How do English and Spanish speakers map this part of the vowel space, and how does this mapping affect their discrimination functions?
Two experiments address these issues. In Experiment I, Swedish, English, and Spanish speakers are asked to label and rate the goodness of an area of the vowel space approximately equivalent to that shown in Figures 1, 2, and 3. In Experiment II, the same speakers perform a discrimination task for a subset of vowels in Experiment I which cross a category boundary for Swedish speakers but not for English or Spanish speakers.

2. EXPERIMENT I: METHODS

2.1. Participants

Fifteen Swedish speakers, fifteen American English speakers, and eleven Spanish speakers participated in this study. The Swedish speakers were run at Stockholm University in Sweden, the English speakers at the University of Washington in Seattle, and the Spanish speakers at the Texas Intensive English Program in Austin. All participants were adult, native speakers of their particular language who reported having no known hearing impairments. All speakers were paid for their participation.

2.2. Apparatus

The stimuli were synthesized at a 10kHz 16-bit sample per second. They were presented by a SoundBlaster 16 digital audio board controlled by a Compaq 486 microcomputer which played through the right-ear speaker of a pair of Telephonics TDH 39P headphones. The participants sat in a quiet room and responses were entered and recorded using the computer that controlled the presentation of the stimuli.

2.3. Stimuli

There were sixty-six stimuli in an F1-F2 grid, all equally spaced along the mel scale. F1 ranged from 250 mels (189 Hz) to 500 mels (414 Hz) and F2 ranged from 1300 mels (1462 Hz) to 1800 mels (2482 Hz). The stimuli were synthesized as monophthongs at 50 mel steps in each dimension. F3 was calculated using a regression formula on F1 and F2 that was published in Neary (1989) and shown to be a good fit for both Swedish and English vowels. The function is as follows: F3 Hz= 0.522(F1 Hz) + 1.197(F2 Hz) + 57. The tokens were equalized in RMS intensity and were 615 ms long. Other synthesis parameters were the same for all tokens.

2.4. Procedure

Identification and goodness proceeded in three stages. In the first stage, participants were asked whether the sound was a vowel in their native language or not. If the answer was yes, then in the second stage the listener identified the vowel with reference to his/her native language categories. The listener did this by referring to a list of common words in that particular language which exemplified all the vowel categories. In the third stage, the listener rated the goodness of the vowel on a scale of 1 (poor) to 7 (good). If in the first stage the participant said that the vowel was not in their language, then he/she was still asked to choose the vowel that best matched what they heard, although they did not have to rate the goodness of that sound. The instructions for the task and the computer interface were all in the native language of the participant. Each participant completed a block of ten practice trials presented in random order. After the practice, they completed an experimental session of 198 trials (3 blocks of 66 tokens) with the order of trials randomized within each block. Participants were allowed to hear the sounds as many times as they needed in order to make their judgments.

2.5. Results

A stimulus had to be reported as belonging to a native language vowel category in stage 1 of the experiment at least 50% of the time in order to undergo identification and goodness calculations. Stimuli that did not reach this criterion are represented by an empty space in Figures 4, 5, and 6. The category identified the majority of the time was assigned to that token. Goodness ratings for the majority category were averaged across subjects and are shown for each token. The identification responses observed in Figures 4, 5, and 6 are similar to the production data in Figures 1, 2, and 3. The main departure from the production data is that English speakers identify part of the front rounded vowel space as the back rounded vowel /u/ as in boot, though they do not give this category a very high goodness rating. Spanish speakers, however, do not assimilate the front rounded vowel to their back rounded vowel /u/. In addition, English speakers did not use the lax /I/ vowel in their labeling. Although the grid encompasses formant frequencies normally associated with /I/, it may be that the duration of the stimuli (615 ms) favors the identification of longer, tense vowels. Goodness ratings of 4 and above for English /i/ and /e/, Spanish /i/ and /e/, and Swedish /i:/, /e:/, /y:/, /ö:/, /ç:/ show that all three language groups found at least one good example of these vowel categories in the grid.
Identification functions appear to be very different across the three languages. If language experience affects vowel discrimination, Swedish speakers should show greater sensitivity than English and Spanish speakers between F1 = 320 and 366 Hz, which straddles the boundary of a Swedish vowel category. Furthermore, English speakers should show greater sensitivity than Spanish speakers due to the fact that they separate /u/ and /er/ at this juncture for these stimuli. To test this, we ran a discrimination experiment on stimuli which differed across languages as to whether they contained a category boundary or not.

3. EXPERIMENT II: METHODS

3.1. Participants, Apparatus, & Stimuli

These are the same as in Experiment I.

3.2. Procedure

The horizontal vector corresponding to F2 = 1828 Hz was selected for the discrimination experiment because Swedish speakers show a category boundary for /ø:/ between F1 = 320 Hz and F1 = 366 Hz and English and Spanish speakers do not. The five pairs of adjacent stimuli along this vector were used in an AX discrimination task with a roving design. On half the trials, participants heard the same stimulus presented twice. On the other half of the trials, participants heard two different stimuli. The offset-to-onset interval was 250 ms for each stimulus pair. After each response, participants received feedback indicating whether their response was correct. Instructions, feedback, and other aspects of the computer interface were in the native language of the participant. Participants first completed a practice session of 5 trials. The experimental session contained 10 blocks of 20 stimuli pairs in random order.
3.3. Results

Sensitivity ($d^\prime$) was measured for each stimulus pair for each subject through the application of signal detection theory (Macmillan & Creelman, 1991). Figure 7 displays the mean discrimination sensitivity for each pair of adjacent stimuli in Swedish, English, and Spanish. Notice that although the stimuli are physically equidistant from each other, sensitivity varies for each pair. A 2 factor ANOVA (Language X Pair) confirmed that Pair was significant beyond the $p < .001$ level. Language was not significant, but there was a significant Language X Pair interaction ($p < .05$). Analysis of the simple effects shows that Language is significant at pairs 4 and 5 ($p < .01$). Also, a planned comparison between the means of the three languages reveals that Spanish is significantly different from English and Swedish ($p < .05$).

There are, however, differences in acoustic sensitivity that phonetic labeling and goodness ratings do not account for. For instance, although they had no phonetic label for the stimuli along the vector, the sensitivity function for Spanish speakers was not monotonic. The peak in sensitivity at pair 3 may reflect a general nonlinearity of the auditory system. Or, it may reflect other aspects of the Spanish vowel system.

4. GENERAL DISCUSSION

The results of this experiment demonstrate an effect of language experience on adults’ perception and discrimination of vowels. Spanish speakers show significantly lower sensitivity to vowel contrasts when compared to Swedish and English speakers, indicating that linguistic experience affects the ability to discriminate small differences in vowels. English speakers assimilated the /æ:/ vowel space to their native /u:/.

That English speakers perceive a category here, albeit not a good one, may explain why the English sensitivity function is not significantly different from the Swedish. Recall that Stevens et al. (1969) also found no significant differences between Swedes and Americans in sensitivity functions. In Stevens et al., however, Americans did not use their native categories in the labeling task, thereby obscuring the relationship between phonetic labeling and discrimination.

There are, however, differences in acoustic sensitivity that phonetic labeling and goodness ratings do not account for. For instance, although they had no phonetic label for the stimuli along the vector, the sensitivity function for Spanish speakers was not monotonic. The peak in sensitivity at pair 3 may reflect a general nonlinearity of the auditory system. Or, it may reflect other aspects of the Spanish vowel system. Figure 6 suggests that the /i:/ - /e:/ boundary for Spanish speakers is at F1 between 275 and 320 Hz. This may have contributed to the peak in sensitivity in Figure 7. Moreover, English speakers may show a peak at pair 4 because this is where Figure 5 suggests their /i:/ - /e:/ is. Similarly, Swedish listeners’ excellent discrimination of pairs 3 and 4 may be explained by their use of F1 to separate the /i:/ - /e:/ and the /æ:/ - /ø:/ contrasts in these respective locations. The fact that the vector varied only in F1 may have focused listeners’ attention on the F1 dimension. On this view, vowel discrimination is not just a function of phonetic labeling, but due instead to the entire vowel map as opposed to a single pair of vowels. Similar thinking can be applied to the trough in sensitivity for all languages at pair 2. This too may be a nonlinearity inherent in the auditory system. Or, it may be that something like a “magnet effect” is produced by a systemic warping of the vowel space which lowers sensitivity in this area of F1 because it lies within the /i:/ category for all three languages.

Several processes may converge to produce vowel labeling and discrimination. Nevertheless, our data suggest a strong relationship between linguistic experience and vowel perception.

5. REFERENCES


