Recognition of a Spanish VV Sequence

J.A. Gurlekian and H.E. Franco
Buenos Aires, Argentina

1. Introduction

The fact that Spanish vowels are accurately identified in isolation and in context seems to present a suitable basis to explore the relevance of some dynamic information cues relative to steady state information cues. In this paper we will attempt to define some interactions of such types of acoustic information which consider listeners' responses to each of the speech-like sequences synthesized.

The Spanish language presents a relatively simple vocalic system with only five categories, while duration is phonologically not distinctive. The five vowels are phonetically monophthongs and they possess a known perceptual stability in running speech.

Recognition of natural and synthetic Spanish vowels in isolation is based on frequency bands sustained along preferred targets. Using mixed speakers, identification scores of 100% were obtained for natural sounds (Guirao and Manrique, 1972) and 100% for synthetic stimuli generated along the F1-F2 plane (Manrique and Gurlekian, 1976). In these experiments no type of training was necessary.

On the other hand changing spectral cues have been claimed to be responsible for high identification scores of vowels in other languages. English listeners in particular seem to improve recognition when this type of information is available. This occurs principally in vowels with consonant context (Strange et al., 1976; Gottfried and Strange, 1980) and it also occurs in diphthongized vowels, (Assman et al., 1982). Strange and Gottfried (1980) obtained the lowest error rate in the identification of the isolated English vowels /o/ and /e/. Although recent experiments have shown that spectral changes, as provided by consonant context, are not an essential condition for vowel identification in the English language, higher identification scores are obtained for all isolated vowels, when listeners' training and task conditions are considered (Macchi, 1980; Diehl, 1981; Assmann, 1982). In a consonant context situation Spanish listeners preserve high identifiability of natural vowels, in spite of the fact that steady state cues are altered, showing that a dynamic spectral pattern, clearly a non-essential cue, may interact with stationary cues like those stated at the isolated emission level.

With this perspective, steady state cues which completely define the Span-
ish vowels in isolation will be presented for identification with additional 
dynamic information defined with relation to a context vowel. In the first 
experiment we expect to determine the effect of the spectral distances to 
different context vowels on the same initial frequency resonances of the 
vowels under study, and the effect of vowel duration, as a variation of the 
steady state cue, on the spectral distance to the context vowel. In the second 
experiment we explore the relevance of the temporal sequence on the identi- 
fication of the vocalic pairs presented in Experiment 1, with the stimuli 
presented backwards. In a third experiment the effect of three different 
constant levels of fundamental frequency is analyzed in terms of the identifi- 
cation scores for the three different tested vowels.

2. Experimental Procedure

The sound /a/ with formant transitions is considered the ‘context’ or affec-
ting vowel, based mainly on the hypothesis of perceptual stability of this 
central sound. This selection allows us a simpler and wider control of the 
formant transitions, resonances and durations of the vowel context in the 
synthesis process than the control required for a consonant environment 
where the perceptual stability of the affecting context might be lost and 
therefore confuse the listeners in their responses to the affected vowel. The 
vocalic segments under study consist of ‘o1’: a good example of an isolated 
vowel /o/ (100% identification) and two vowels used as control ‘02’ (80% 
identification as /o/) and ‘o3’ (50% identification as /o/) obtained from a 
/o-a/o/ Spanish vocalic continuum presented for identification in a free 
choice condition (Manrique and Gurlekian, 1976). Two /a/ vowels are 
synthesized, both with formant values taken from the same continuum. 
These values can be considered as those produced in isolation by two 
different vocal tract lengths, one is called ‘normal’ /a/ or a(n), and the 
second corresponding to a shorter vocal tract (viz. higher resonances) is 
called ‘elevated’ /a/ or a(e). The vocalic segments are all combined as /o-a/ 
sequences through linear interpolation. Also, for each combination, we 
reverse the duration of the first vocalic segment systematically, thus imitating 
a speaking rate effect as it occurs in running speech.

In Experiment 1 the stimulus variations are: (1) duration of the first vocalic 
segment which ranged from 7.6 msec to 110 msec in steps of 25.6 msec, and 
(2) formant values, indicated in Fig. 1. The fundamental frequency was held 
constant at 130 Hz. The duration of the transition portion was also held 
constant at 70 msec for all the stimuli and the duration of the second vocalic 
segment was fixed to 110 msec. The stimuli were generated by a formant 
synthesizer designed by Klatt (1981). Three test tapes were generated. Each 
of them consisted basically of two VV sequences. One was made with a 
particular /o/ structure and the ‘normal’ /a/. The second VV sequence was 
composed of the same /o/ structure and the ‘elevated’ /a/. From these two 
sequences, five new stimuli were generated through digital segmentation of 
the first vocalic segment. This made a total of 10 stimuli which were then 
repeated at random five times for each identification test. Interstimulus 
intervals were always 5 sec. Ten native speakers, graduate students of the 
University of Buenos Aires, were asked to identify each sequence in the tests 
as the combination of two Spanish vowels.

In Experiment 2 the synthetic waveforms of Exp. 1a, 1b, 1c were digitally 
reversed, and the same identification was requested.

In Experiment 3, the stimuli consisted of three basic VV sequences made 
with the three /o/ versions and the normal /a/ only. Each of these sequences 
have alternatively constant fundamental frequencies of 130 Hz, 230 Hz and 
330 Hz. Three values of the duration of the vocalic segment /o-a/ was tested 
in different sessions. Again the subjects were asked to identify the stimuli as 
the concatenation of two Spanish vowels.

Both /oa/ and /ua/ are meaningless sequences in Spanish and they appear 
as syllables with similar frequencies of occurrence. Also, when they appear in 
CVV sequences before /a/, /o/ and /a/ seem to lose their oppositional 
character and to remain in free variation.

3. Results

The results of the experiments are presented as identification scores of the 
/oa/ sequence in every case. For each of the /o/ vowels tested in Exp. 1a, 1b 
and 1c there are two identification functions in Fig. 2, showing the effect of 
spectral distances (different slopes) to a(n) or to a(e) upon different dura-
tions of the /o/ segment. Higher identification as /u/ is observed when greater spectral distances occur between the vocalic segments in the identification functions, as shown in Fig. 2. Identification scores for o1, o2 and o3 at the maximum duration tested closely correspond to the percentages obtained in the isolated identification of this same segment in the previous work of Marique and Gurlekian (1976). At shorter durations of the /o-u/ segment, identification moves progressively to /u/.

These results indicate that: (1) recognition of a vowel like /o/ in isolation can be changed towards /u/ when affected by the spectral distance to a context vowel, despite the fact that formant values and the duration of the vocalic segment remain unchanged; (2) as a second and stronger effect, increasing the influence of the context vowel and the associated transitions, shorter durations reduce the information of the steady state cue reciprocally. Trading relations can be established for both a higher spectral distance and a shorter duration of the vocalic segment and formant frequencies of the stimuli.

Results from Experiment 2 (Fig. 3) show lower percentages of identification as /o/ for o1 and o2 versions than those obtained in Exp. 1. This is an indication that the temporal distribution of the acoustic cues: resonance frequencies of /a/ plus transitions that can be considered adequate to /u/ in running speech, and conflicting resonance frequencies for /o/ or /u/ may interact asymmetrically according to responses in Exp. 1 and Exp. 2 where all spectral and temporal parameters are the same. We have hypothesized that a preassignment occurs in the perceptual process as the listeners find some typical speech pattern and this first decision is highly responsible for our final phonological assignment. In Exp. 2 a clear acoustic pattern of /au/ is available at first, then a conflicting pattern of /u-o/ appears that diminishes the listeners' final responses. In the Spanish vowel system, areas of /u/ and /o/ are aligned in a radial direction in the F1-F2 space. Due to the proximity of /u/ and /o/ areas, a partial overlapping occurs between them when considering emissions from
Table I. Percent of /oa/ responses

<table>
<thead>
<tr>
<th>Duration of the /o-u/ segment</th>
<th>F₀ (msec)</th>
<th>Vowel version (Hz)</th>
<th>o₁</th>
<th>o₂</th>
<th>o₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>110</td>
<td>130</td>
<td>100</td>
<td>77.5</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>230</td>
<td>100</td>
<td>42.5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>330</td>
<td>55</td>
<td>20</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>58.8</td>
<td>130</td>
<td>97.5</td>
<td>77.5</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>230</td>
<td>77.5</td>
<td>12.5</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>330</td>
<td>57.5</td>
<td>12.5</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>7.6</td>
<td>130</td>
<td>52.5</td>
<td>27.5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>230</td>
<td>42.5</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>330</td>
<td>27.5</td>
<td>2.5</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

different speakers. Since there is a correlation between the fundamental frequency ranges and the different vocal tract lengths for children, women and men, the results suggest that listeners have used F₀ (in the absence of higher formant information) to assign the same formant pattern to /o/ or /u/.

References


