EFFECT OF PHONETIC CONTEXT ON WOMEN’S VOWEL AREA

Megan M. Hodge, Janine Chesworth, Kim Coté, Colleen Shaw, Sandra Williams

Department of Speech Pathology & Audiology, University of Alberta
Edmonton, Alberta, Canada
megan.hodge@ualberta.ca

ABSTRACT
Vowel quadrilateral area (VQA) refers to the area enclosed in the quadrilateral formed when the coordinates for the first and second formant frequencies (F1 and F2) of the corner vowels (/i/, /æ/, /a/ and /u/) are plotted. VQA is correlated positively with speech intelligibility scores. This study examined how size and shape of VQA is affected by phonetic context for 12 women with typical speech production. Four phonetic contexts were compared: /hV/, /hVd/, a subset of 24 words from a children’s speech intelligibility test, and 6 sets of minimally contrastive CVC words. Six tokens of each vowel in each context were recorded from each talker. Only vowel tokens identified as members of the target vowel category were included in the VQA calculations. F1 and F2 frequencies of the four corner vowels were estimated from wideband spectrographic displays of the talkers’ digital audio recordings. VQAs were calculated in log Hz for each talker in each phonetic context condition. As predicted, VQA was affected significantly by phonetic context. The largest area was obtained in the most phonetically neutral (/hV/) condition and the smallest areas were obtained in the least neutral conditions. Change in size and shape of the vowel quadrilateral from most to least neutral contexts reflected reduction in both F1 and F2 ranges but was significant only for F2. Inspection of the vowel quadrilateral plots revealed that increases in F2 frequency values for /u/ contributed most to reduced VQA in non-neutral phonetic contexts.

INTRODUCTION
This study addressed a basic methodological issue for a variable that has been identified as one important acoustic correlate of speech intelligibility. This variable is “vowel quadrilateral area” or VQA. VQA refers to the area enclosed in the quadrilateral formed by a plot of the coordinates for the first and second major vocal tract resonances (formant frequencies) of the corner vowels (/i/, /æ/, /a/ and /u/). Previous studies have used VQA as an acoustic correlate of human perceptual judgments of speech intelligibility. Because VQA has been found to account for between 41% and 53% of the variance observed in speech intelligibility scores (Higgins & Hodge, 2002; Hodge, 1999; Weismer et al., 2001), it has been identified as a variable of interest in acoustic modeling of speech intelligibility. One methodological concern when comparing VQA across studies of similar speakers, and across speakers who differ in stage of speech development, is the phonetic context in which the corner vowels are embedded. It is known that sounds surrounding a vowel influence the vowel’s first (F1) and second (F2) formant frequencies (Hillenbrand et al., 2001; Stevens & House, 1963). Therefore differences in VQA reported by previous investigators are difficult to compare and interpret because these differences may be due to phonetic context effects as opposed to, or in addition to, speaker differences (e.g., dialect, speaking style, age).
This study addressed the following questions:

1) Is there an effect of phonetic context on the size of F1 by F2 VQA? Based on previous research, it was hypothesized that VQAs would be smaller when the target vowels occurred in less neutral phonetic contexts. This question was addressed by calculating and comparing F1 by F2 VQAs using a natural log Hz scale for each of four contexts. The log Hz scale was used to reduce the effect of interspeaker vocal tract size differences on vowel formant frequency values (Nearey, 1992).

2) Is there an effect of phonetic context on the shape of F1 by F2 VQA? This question was investigated by comparing the size of F1 and F2 ranges between the most and least neutral phonetic context. It was hypothesized that F1 and F2 ranges for the vowel quadrilateral would be smaller in less neutral phonetic contexts.

METHOD

Phonetic Context Conditions
Monosyllabic utterances in four phonetic contexts were compared: /hV/, /hVd/, a subset of 24 words from the Test of Children’s Speech or TOCS (children’s speech intelligibility test) (Higgins & Hodge, 2002), and 6 sets of minimally contrastive CVC words, referred to as the MC condition. The stimulus words for the TOCS and MC conditions are listed in Appendix A. The /hV/ context was categorized as the most phonetically neutral since no supraglottal articulation is used for /h/. The MC context was categorized as the least phonetically neutral because all target monosyllables had both an initial and final consonant.

Speakers
Twelve women between the ages of 20 and 35 who met the following inclusion criteria provided vowel recordings for analysis: 1) no history of speech disorder or treatment; 2) Western Canadian dialect of English as their first language; 3) normal hearing demonstrated by passing a standard hearing screening; 4) non-smokers for the past seven years; and 5) free from colds, sore throats, or any adverse health condition that may have affected their voice at the time of recording.

Vowel Recording
Digital audio recordings (sampling rate of 48 kHz, 16 bit quantization) of each speaker’s vowel productions for each phonetic context were made in a quiet environment. Speakers wore a professional, unidirectional head-mounted microphone placed 1.5 inches away from the corner of the mouth. Speakers were familiarized with pronunciation of the /hV/ and /hVd/ printed stimuli prior to recording. Each stimulus word was presented on a computer screen. Eight practice words were presented to familiarize speakers with the task. Speakers then produced the 96 test items (24 items per condition, times 4 conditions) in random order. A panel of three judges provided independent verification that the vowel token in each recording was a perceptually valid member of the target vowel category.
Formant Frequency Measurement
The F1 and F2 values of the target vowels in each speaker’s digital audio files were estimated from wideband spectrograms using CSpeech 4.0 (Milenkovic, 1992). The analyzing bandwidth was at least twice the speaker’s highest fundamental frequency (F0) for /i/ and ranged between 450 and 550 Hz. To isolate the vowel, the cursors were placed on the first and last glottal pulses that excited the first two formants of the vowel segment. The length of the vowel was measured as the distance between the cursors. F1 and F2 values were estimated visually from a 30-millisecond window that was centered at the midpoint of the vowel segment using FFT and LPC power spectra as guides.

Calculation of Vowel Quadrilateral Areas
Mean F1 and F2 frequencies were calculated for the six tokens for each vowel category in each condition using a log Hz scale. For example, F1 values for a speaker’s six TOCS /i/ vowel tokens were averaged to give a mean F1 value. This was repeated for F2 values. These mean F1, F2 coordinates for each corner vowel were used to generate vowel quadrilaterals for each speaker in each phonetic context using a log Hz scale. VQA was calculated using the formula reported by Higgins and Hodge (2002).

\[
VQA = 0.5 \times \{(i/F2 \times /æ/F1 + /æ/F2 \times /a/F1 + /a/F2 \times /u/F1 + /u/F2 \times /i/F1) \\
(i/F1 \times /æ/F2 + /æ/F1 \times /a/F2 + /a/F1 \times /u/F2 + /u/F1 \times /i/F2)\}
\]

Calculation of and F1 and F2 Ranges
A formant range was defined as the difference in log Hz between the smallest and largest formant value, for a given formant, in a given phonetic context condition, for each speaker. For F1, this was the range covered by the vowel quadrilateral plot on the x axis and for F2, the range of the vowel quadrilateral plot on the y axis.

RESULTS
Group mean VQAs for each condition are shown in Table 1. As predicted, VQAs were greatest for the neutral context /hV/ and smallest for the two non-neutral contexts (TOCS and MC words). A repeated measures 1-way ANOVA revealed a significant effect of phonetic context on VQA \((F(3, 33) = 24.2, p < .001)\).

Group mean F1 and F2 ranges for each condition are shown in Tables 2 and 3. Statistical testing using 1-way ANOVAS with repeated measures revealed no significant difference across phonetic contexts for F1 \((F(3, 33) = 1.5, p = .226)\) but a significant difference for F2 \((F(3, 33) = 15.6, p < .001)\).
Table 1. Mean VQA for each phonetic context.

<table>
<thead>
<tr>
<th></th>
<th>Mean (log Hz²)</th>
<th>Std. Deviation (log Hz²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/hV/</td>
<td>.587</td>
<td>.149</td>
</tr>
<tr>
<td>/hVd/</td>
<td>.538</td>
<td>.129</td>
</tr>
<tr>
<td>TOCS</td>
<td>.473</td>
<td>.120</td>
</tr>
<tr>
<td>MC</td>
<td>.450</td>
<td>.146</td>
</tr>
</tbody>
</table>

Table 2. Mean F1 ranges for each phonetic context.

<table>
<thead>
<tr>
<th></th>
<th>Mean (log Hz)</th>
<th>Std. Deviation (log Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/hV/</td>
<td>1.024</td>
<td>.134</td>
</tr>
<tr>
<td>/hVd/</td>
<td>1.000</td>
<td>.146</td>
</tr>
<tr>
<td>TOCS</td>
<td>1.009</td>
<td>.143</td>
</tr>
<tr>
<td>MC</td>
<td>.973</td>
<td>.150</td>
</tr>
</tbody>
</table>

Table 3. Mean F2 ranges for each phonetic context.

<table>
<thead>
<tr>
<th></th>
<th>Mean (log Hz)</th>
<th>Std. Deviation (log Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/hV/</td>
<td>.922</td>
<td>.143</td>
</tr>
<tr>
<td>/hVd/</td>
<td>.860</td>
<td>.130</td>
</tr>
<tr>
<td>TOCS</td>
<td>.797</td>
<td>.087</td>
</tr>
<tr>
<td>MC</td>
<td>.776</td>
<td>.082</td>
</tr>
</tbody>
</table>

DISCUSSION AND CONCLUSIONS

As predicted, VQA was affected significantly by phonetic context. The largest area was obtained in the most phonetically neutral (/hV/) condition and the smallest areas were obtained in the least neutral (TOCS and MC) conditions. These results are in agreement with those described by Stevens and House (1963). They reported F1 and F2 values for a range of vowels including the corner vowels /i/, /æ/, /a/ and /u/, in three different phonetic contexts (/V/, /hVd/, and /hæCVC/), produced in isolation by three men. VQAs calculated from the F1 and F2 means for
the corner vowels reported by Stevens and House are .655 Hz², .628 Hz², and .437 Hz², for the respective phonetic contexts, with the smallest area corresponding to the least neutral context.

In the current study, change in size and shape of the vowel quadrilateral from most to least neutral contexts was affected by reduction in both F1 and F2 ranges but was significant only for F2. Visual inspection of the vowel quadrilateral plots revealed that increases in F2 frequency values for /u/ contributed most to reduced VQA in non-neutral phonetic context conditions, which is also apparent in the formant measures reported by Stevens and House (1963). Future research will test hypotheses about interactions between age and phonetic context on vowel quadrilateral size and shape and their relationship to speech intelligibility measures.

ACKNOWLEDGMENTS
This project was supported by a grant from the Canadian Language and Literacy Research Network.

REFERENCES


Milenkovic, P., & Read, C. (1992). CSpeech 4.0, Laboratory version [Computer Program], Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI.


### APPENDIX A: TOCS and MC Stimulus Words

<table>
<thead>
<tr>
<th></th>
<th>/i/</th>
<th>/æ/</th>
<th>/a/</th>
<th>/u/</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOCS</td>
<td>heat</td>
<td>hat</td>
<td>hot</td>
<td>hoot</td>
</tr>
<tr>
<td>Words</td>
<td>“D”</td>
<td>badge</td>
<td>Hawk</td>
<td>Chew</td>
</tr>
<tr>
<td>eat</td>
<td>hash</td>
<td>Chop</td>
<td>Shoe</td>
<td></td>
</tr>
<tr>
<td>beat</td>
<td>bad</td>
<td>Jaw</td>
<td>Sue</td>
<td></td>
</tr>
<tr>
<td>seat</td>
<td>hatch</td>
<td>Stop</td>
<td>Two</td>
<td></td>
</tr>
<tr>
<td>sheet</td>
<td>pad</td>
<td>top</td>
<td>zoo</td>
<td></td>
</tr>
<tr>
<td>MC</td>
<td>seat</td>
<td>sat</td>
<td>sought</td>
<td>suit</td>
</tr>
<tr>
<td>Words</td>
<td>beat</td>
<td>bat</td>
<td>bought</td>
<td>boot</td>
</tr>
<tr>
<td>G’s</td>
<td>jazz</td>
<td>jaws</td>
<td>Jews</td>
<td></td>
</tr>
<tr>
<td>keep</td>
<td>cap</td>
<td>cop</td>
<td>coop</td>
<td></td>
</tr>
<tr>
<td>teak</td>
<td>tack</td>
<td>talk</td>
<td>toque</td>
<td></td>
</tr>
<tr>
<td>heat</td>
<td>hat</td>
<td>hot</td>
<td>hoot</td>
<td></td>
</tr>
</tbody>
</table>