EFFECT OF PHONETIC CONTEXT ON WOMEN'S VOWEL AREA

Janine Chesworth, Kim Coté, Colleen Shaw, Sandra Williams & Megan Hodge
Department of Speech Pathology & Audiology, University of Alberta, Rm. 2-70 Corbett Hall, Edmonton, AB, T6G2G4
(Project of the Canadian Language and Literacy Research Network)

1. 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[1,2] 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 and second formant frequencies[3,4], abbreviated as F1 and F2. To our knowledge no study has been published that has directly investigated how various surrounding speech sounds, that is, phonetic context, influence the size of VQA in the same sample of speakers. 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 the 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[5].

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 extents between the most and least neutral phonetic context. It was hypothesized that F1 and F2 extents for the vowel quadrilateral would be smaller in less neutral phonetic contexts.

2.METHOD

2.1 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)[2], 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.

2.2 Speakers

Ten 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.

2.3 Vowel Recording

Digital audio recordings (sampling rate = 44 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 randomized order. A panel of three judges provided independent verification that the vowel in each recording was a perceptually valid member of the target vowel category.

2.4 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[6]. The analyzing bandwidth was at least twice the speaker's highest fundamental frequency (F0) for /i/ and was in the range of 450 to 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.

2.1 Calculation: Vowel Quadrilateral Areas

Mean F1 and F2 frequencies were calculated for the six tokens for each vowel category in each condition, in the 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[2].

2.2 Calculation: Formant Extents

A formant extent 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 would be 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.

3. RESULTS

3.1 Vowel Quadrilateral Areas

Group mean VQAs for each condition are shown in Table 1. As predicted, VQAs are 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 ($F(3,27) = 33.0$, $p = .0001$). Results of post-hoc testing indicated that all conditions differed significantly from each other ($p < .008$ – Bonferroni correction) except for the *TOCS* and MC conditions ($p = .015$).

Table 1. Mean VQA for each phonetic context.

	Mean (log Hz^2)	Std. Deviation (log Hz^2)
/hV/	.544	.122
/hVd/	.502	.109
TOCS	.442	.103
MC	.401	.862

3.2 F1 and F2 Extents

Group mean F1 and F2 extents for each condition are shown in Table 2. Statistical testing using 1-way ANOVAS with repeated measures revealed no significant difference across phonetic contexts for F1 ($F(3,27)$ $p = .129$) but a significant difference for F2 ($F(3,27)$ $p < .000$). Post-hoc testing revealed significant differences between the /hV/ and *TOCS* conditions and between the /hV/ and MC conditions ($p<.008$ – Bonferroni correction).

Table 2. Mean F1 and F2 extents for each phonetic context.

	/hV/	/hVd/	*TOCS*	MC
F1 Extent				
Mean (log Hz.)	.992	.951	.959	.931
Std. Dev.	.121	.102	.090	.108
F2 Extent				
Mean (log Hz.)	.889	.835	.787	.779
Std. Dev.	.133	.128	.088	.067

4. DISCUSSION AND CONCLUSIONS

As predicted[3,4], 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. Change in shape of the vowel quadrilateral from most to least neutral contexts was affected by reduction in both F1 and F2 extents but was significant only for F2. Visual inspection of the VQA plots (not shown here) revealed that increases in F2 frequency values for /u/ contributed most to reduced VQA in non-neutral phonetic context conditions. Future research will test hypotheses about interactions between age and phonetic context on vowel quadrilateral size and shape.

5. REFERENCES

1. Weismer, G., Jeng, J., Laures, J., Kent, R., & Kent, J. (2001). Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. Folia Phoniat. et Logopaedica. 53, 1-18.
2. Higgins, C. & Hodge, M. (*2002b*). Vowel area and intelligibility in children with and without dysarthria. J. Medical Speech-Language Pathology, 10(4), 271-274.
3. Hillenbrand, J., Clark, M. & Nearey, T. (2001). Effects of consonant environment on vowel formant patterns. J. Acoust. Soc. Am., 109 (2), 748-763.
4. Stevens, K. & House, A. (1965). Perturbation of vowel articulations by consonantal context: An acoustical study. J. Speech Hear. Res.6, 111-121.
5. Nearey, T. (1992). Applications of generalized linear modeling to vowel data, Proceedings, 1992 International Conference on Spoken Language Processing, 583-586.
6. Milenkovic, P., & Read, C. (1992). CSpeech 4.0, Laboratory version [Computer Program], Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI.

6. APPENDIX A: *TOCS* and MC Stimulus Words

	/i/	/ae/	/a/	/u/
TOCS				
	heat	hat	hot	hoot
	"D"	badge	hawk	chew
	eat	hash	chop	shoe
	beat	bad	jaw	sue
	seat	hatch	stop	two
	sheet	pad	top	zoo
MC				
	seat	sat	sought	suit
	beat	bat	bought	boot
	G's	jazz	jaws	Jews
	keep	cap	cop	coop
	teak	tack	talk	touque
	heat	hat	hot	hoot