PRELIMINARY STATISTICAL PATTERN RECOGNITION METHODS IN THE STUDY OF VOWELS PRODUCED BY CHILDREN WITH AND WITHOUT SPEECH SOUND DISORDERS

Hyunju Chung[1], Terrance, M. Nearey[2], Megan Hodge[1], Karen, E. Pollock[1], and Benjamin, V. Tucker[2]
[1]Dept. of Speech Pathology & Audiology, University of Alberta, 8205 114St., Edmonton, AB, Canada, T6G 2G4
[2]Dept. of Linguistics, 4-32 Assiniboia Hall, Edmonton, AB, Canada, T6G 2E7

1. INTRODUCTION

The nature of vowel acquisition and vowel error patterns in young children, especially those with speech sound disorders (SSD), is not well understood. Previous studies of vowels produced by children with SSD of unknown origin (SSD-UNK) have shown that they demonstrate a similar order of acquisition and vowel error patterns as those of children with TDS (typically developing speech), but are slower in developing vowels and produce more vowel errors (e.g., Stoel-Gammon & Herrington, 1990). Previous studies have relied mainly on phonetic transcription to describe the characteristics of vowels produced by children with SSD. A few studies used acoustic analysis to investigate the vowels of children with SSD due to motor impairments. These studies have shown that children with cerebral palsy have smaller vowel space size than children with TDS, and that size of vowel space is correlated positively with their speech intelligibility (e.g., Higgins & Hodge, 2001). These studies, however, are based on the acoustic measurements of only 3 to 4 corner vowels. In addition, despite the proposed importance of spectral movement patterns in successful identification of vowels (e.g., Hillenbrand, & Nearey, 1999), only a few studies (e.g., Lee, 2009) have examined spectral movement patterns of vowels produced by young children with SSD.

In the current study, acoustic characteristics of each of the 10 English monophthong vowels were examined and compared to the same-aged children with and without SSD. For acoustic measurements, F1, F2, and fundamental frequency (F0), that incorporated spectral movement patterns, were obtained for the analysis. Using these measurements, statistical pattern recognition models (e.g., Thomson, Nearey, & Derwing, 2009) were used to examine whether sets of acoustic variables differentiate 1) vowels produced by children with and without SSD, 2) age subgroups within each child group, and 3) vowels that are identified accurately from those that are not.

2. METHOD

2.1 Participants

Adult participants were 15 women, ages 18 to 35 years. All were monolingual speakers of Western Canadian English with no history of speech delay or disorder. Two groups of children participated (TDS and SSD-UNK), with 3 children in each of four age groups (3, 4, 5, and 6-year-olds) in each group for a total of 24 children. All children were learning English as their first language and living in Western Canada. Parents of children in the TDS group reported no concerns about their child's speech and language development, and all children passed a standard speech and language screening measure. Children in the SSD group were receiving or on waiting lists for speech therapy.

2.2 Stimuli

The target words used in this study are a subset of words from the three *TOCS+* word lists (Hodge, Daniels, & Gotzke, 2009) (Table 1). Target vowels were 10 English monophthongs, [i, ɪ, e, ɛ, æ, ʌ, ɑ, o, ʊ, u].

Table 1. List of target words and vowels

Vowel	Target Words	Vowel	Target Words
/i/	bead, beat, bee, D, feet, tea, peep	/ɪ/	bit, fit, hid, sit
/e/	bait, pain	/ɛ/	bet, pen
/æ/	baa, bad, bag, fat, hat, pat, tap	/ʌ/	bud, bug, hut, pup, shut, tub
/ɑ/	Don, hot, jaw, paw, pop, pot, shot, top	/o/	cone, toe
/ʊ/	foot, hood, soot	/u/	boo, Pooh, shoot, hoot, suit, two, tube

2.3 Analysis

The boundaries for each vowel token were manually defined using *Praat* (Boersma & Weenink, 2012). A semi-automatic formant tracking program (Nearey *et al*., 2002) created in MATLAB (7.8.0.347, R2009a) extracted vowel duration: it also extracted F1, F2, and F0 at 2ms steps over the entire duration of the vowel. For each vowel, the last 10ms or the earliest point where the amplitude falls 25dB below the peak has been trimmed. All F1 and F2 values were then log transformed. For F0, the median of the first half of the trimmed vowel was used. The log transformed F1 and F2 measured at 20% and 70% time points, median F0, and duration were used as input variables for the pattern recognition model. The acoustic measures of vowels produced by all speakers were used to train a linear discriminant analysis model. Predicted identification rates for each group of speakers were calculated using the resubstitution method that is the same data was used in training the model and predicting the classification.

3. RESULTS

The preliminary analyses indicated that adult vowels were classified with the highest accuracy (91.7%). Vowels of two groups of children were classified with similar accuracy (TDS - 80.6% and SSD-UNK - 74.9%), but at lower accuracy than for the adult vowels. Across age groups, vowels of the 6-year-olds were classified with higher accuracy than those of the younger age groups in each of the TDS and SSD-UNK group. Across all groups, /i/ was classified most accurately and the vowels /ɛ/ or /ʌ/ least accurately. The model predicted accuracies of 10 vowels of each speaker group are summarized in Table 2.

Table 2. Overall classification accuracy (%) for vowels produced by each speaker and age group.

Speaker Group	i	ɪ	e	ɛ	æ	ʌ	ɑ	o	ʊ	u
	Vowel Category									
Adults	98	100	84	63	88	84	95	100	92	96
TDS 6yr	100	100	100	100	63	82	86	100	83	86
TDS 5yr	100	100	25	50	86	62	79	75	50	100
TDS 4yr	100	100	50	25	87	54	69	100	60	86
TDS 3yr	100	100	100	0	92	50	75	50	50	93
SSD-UNK 6yr	100	88	100	60	64	54	69	100	80	93
SSD-UNK 5yr	87	88	25	0	92	75	79	50	40	86
SSD-UNK 4yr	75	100	100	100	100	39	73	100	17	87
SSD-UNK 3yr	100	100	50	33	80	36	64	50	60	54

Vowels were better identified with two measurements (at 20% and 70%) of the vowel formant pattern than a single measurement (at 50%) for all groups. Classification scores were higher when all the acoustic variables were entered than when either duration or F0 was absent (Table 3). The result of the two-way ANOVA showed that each or the combination of acoustic measurements differs significantly by vowel type and speaker groups.

Table 3. Overall classification accuracy (%) of each speaker group by acoustic variables.

Speaker Group	Acoustic measures: duration F0		no duration F0		duration no F0		no duration no F0	
	20,70%	50%	20,70%	50%	20,70%	50%	20,70%	50%
Adults	91.7	85.5	87.3	83.4	91.1	85.4	85.9	81.9
TDS	80.6	77.3	79.0	73.5	80.6	77.3	78.7	73.0
SSD-UNK	74.9	67.5	73.3	68.8	73.9	66.1	72.5	67.2

4. DISCUSSION

The analysis using a pattern recognition model showed that vowels of adults were better identified than those of the child groups, as expected. The classification accuracy of the two child groups was not very different; accuracy of SSD-UNK was slightly lower than those of TDS group. An age difference in classification accuracy was also found between the oldest children (6-year-olds) and the younger ages. Vowels of all groups were more successfully classified with two measurements representing the formant movement patterns, than a single point measurement, and a combination of all acoustic measures than a single or a subset of measures. Regardless, some vowel categories were always classified with higher accuracy than others.

Our next steps include 1) addressing ways to minimize measurement errors (e.g., rechecking formant frequencies and F0 of poorly classified tokens) and 2) developing methods to compare the model predicted accuracy with the judged accuracy of vowels based on listener identification scores. Further testing of the model will follow, using the same acoustic measurements from additional children with and without SSD.

REFERENCES

Boersma, P. & Weenink, D.(2012). Praat: doing phonetics by computer [Computer program]. Version 5.03.02, retrieved 27 March 2012 from http://www.praat.org/.

Hillenbrand, J.M., & Nearey, T.M. (1999). Identification of resynthesized /hVd/ syllables: Effects of formant contour, *Journal of the Acoustical Society of America, 105 (6)*, 3509-3523.

Higgins, C. M., & Hodge, M. M. (2001). F2/F1 vowel quadrilateral area in young children with and without dysarthria. *Canadian Acoustics - Acoustique Canadienne, 29*(3), 66-67.

Hodge, M., Daniels, J., & Gotzke, C. L. (2009). TOCS+ *intelligibility measures* (version 5.3)[computer software]. Edmonton, AB: University of Alberta.

Lee, J. (2009). *Development of vowels and their relationship with speech intelligibility in children with cerebral palsy.* Unpublished doctoral dissertation, University of Wisconsin, Madison.

MATLAB (2010). Version 7.10.0. Natick, Massachusetts: The MathWorks Inc.

Nearey, T. M., Assmann, P. F. & Hillenbrand, J. M. (2002). Evaluation of a strategy for automatic formant tracking. *Journal of the Acoustical Society of America, 112,* 2323.

Stoel-Gammon, C. & Herrington, P. (1990). Vowel systems of normally developing and phonologically disordered children. *Clinical Linguistics and Phonetics, 4*, 145-160.

Thomson, R. I., Nearey, T. M., & Derwing, T. M. (2009). A modified statistical pattern recognition approach to measuring the crosslinguistic similarity of Mandarin and English vowels. *The Journal of the Acoustical Society of America, 126 (3)*, 1447-1460.