Measurement of child speech complexity using acoustic landmark detection

Marisha Speights[1], Suzanne Boyce[1], Joel MacAuslan[2],Harriet Fell[3] JoHannah Ungruhe [1], Keiko Ishikawa [1]

1. Communication Sciences and Disorders, University of Cincinnati, Cincinnati, OH., 2. Speech Technology and Applied Research, Boston, MA. 3. College of Computer Information Science, Northeastern University, Boston, MA.

Introduction

An important measure of intelligibility in young children is the ability to articulate complex syllables[1-4]. The development of well-formed syllables in infancy has been shown to be a significant predictor of later communication skills. [1-4] Children with delayed speech acquisition do not show this same developmental trend, and deviations in syllable acquisition may serve as a diagnostic marker of future speech delay.[5,6] Historically, syllable shape analysis has required phonetic transcription of speech data which is both training- and time-intensive.[5-7] Researchers have called for automatized methods to quantify measures of articulatory complexity to address this limitation. In this pilot research we test the ability of one such method, Automatic Syllabic Cluster Analysis, to identify differences in syllabic complexity in children who are typically developing and those with diagnosed speech disorders. Automatic Syllabic Cluster Analysis is compared to a conventional hand measure of speech complexity, the Word Complexity Measure (WCM).[15]

SpeechMark Automatic Syllabic Cluster Analysis

The SpeechMark ® Landmark Analysis System locates areas in the acoustic signal that identify where rapid acoustic changes are occurring. Abrupt changes are used by listeners to make perceptual decisions about the speech uttered and are associated with distinctive features.[8-10] The algorithm further groups landmarks into syllabic clusters. These clusters of landmarks represent acoustic patterns that correlate to syllable patterns of English. Statistics derived from these groupings are used to determine the complexity of utterances. Automatic Syllabic Cluster Analysis does not require transcription for analysis because it is not lexically driven. Acoustic parameters measured characterize articulatory precision based upon how the speech was uttered. Different patterns of landmarks will be detected dependent upon the combination of speech sounds as well as how the string of syllables was spoken. Syllables whose production hews more closely to the canonical form will show a characteristic pattern of landmarks. Fewer landmarks may be detected when the same syllables are spoken with less precision in articulatory movements and timing, as commonly found in young children and those with speech-language disorders[10-11]. In past work, syllabic cluster analysis identified significant differences in infant babble when syllable complexity was examined in normal infants and infants at risk for communication disorders.[13-14]

SpeechMark Landmark Analysis System

Based on Liu (1995) & Stevens (1992) detects and clusters three types of landmarks:

1. g: glottis. Marks the time when the vocal folds transition from not vibrating to freely vibrating (+g) or cessation of vibration (-g).

2. s: syllabicity. Marks sonorant and consonantal releases (+s) and closures (-s). These are always voiced.

3. b: bursts. Designates frication onsets or affricates or stop bursts (+ b) and points where aspiration or frication ends (-b) due to stop closure. (Results from simultaneous abrupt changes in frequency bands.) These are never voiced.

Syllabic Clusters Analysis

Landmarks can be grouped into clusters corresponding to syllabic units (syllabic complexity) [10,11,13,14].

V units correspond with a +g, -g sequence

CV units such as “see” when precisely articulated may show up as +b, +g, -g

CVC unit such as “bear” may appear as +b, -b, +g, -s,-g

Landmark patterns reflect syllables as uttered. Acoustical patterns may be weaker and thus landmarks less likely to be detected when speech production is less canonical

Research Question:

1) Can the Syllabic Clusters per Utterance measure predict speaker group (typical vs. disordered)?
2) Does the automated syllabic cluster analysis correlate with a conventional measure of syllabic complexity that uses hand measurement?

Data Collection & Analysis Methodology

Subjects:

6 children - 3 diagnosed as typically developing (age 3-5),
3 diagnosed with speech-language disorder (age 4-5)

Data Collection

- 33 utterances were elicited using a child story book, *Brown bear, brown bear, what do you see?*[16]
- Utterances were recorded using a Shure Wireless Microphone and digitized at 22K

Data Analysis

- Utterance Recordings were analyzed by the Syllabic Cluster algorithm from the SpeechMark Landmark Analysis System®MatLab tool box.
- Measures extracted were: Total Number LMs, LMs per SCs, Number of Utterances, SCs per Utterance
- 99/99 tokens were analyzed for the typical group and 76/99 for disordered group.
- Each token was also phonetically transcribed and scored using the WORD COMPLEXITY MEASURE (WCM).[15] In this conventional analysis, each utterance is scored across eight parameters according in terms of word patterns, syllable structures, and sound classes to measure the complexity of each word.

Landmark Syllabic Clusters

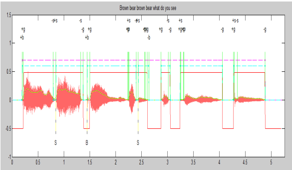


Fig. 1 Normal age 4

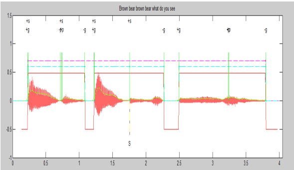


Fig. 2 Speech Sound Disorder Age 5

Landmarks for one utterance of child speech. The LMs are placed at points when abrupt change of energy is occurring simultaneously across multiple frequency ranges at multiple time scales. Waveform with smoothed amplitude envelop, landmarks, and landmarks generated by SpeechMark® MATLAB Toolbox. Solid red line shows the interval of voicing. The dashed blue line indicates the grouping of landmarks into a syllabic cluster. The dashed magenta line shows the grouping for the utterance.

Mean syllabic clusters per utterance by subject.

This figure shows the mean syllabic clusters (SCs) per utterances (Utt) obtained for all tokens produced by the 6 speakers. SCs/Utt is derived from the grouping of landmarks into syllabic clusters per each utterance The average is automatically computed for all utterances analyzed as a measure of the complexity of the entire speech sample.

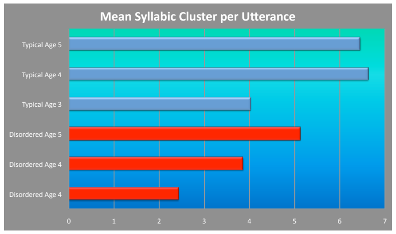


Fig 3.

This work was funded by United States National Institutes of Health (NIH)grants R43 and R44 DC010104, R42 AG033523, and R42 HD34686 to S.T.A.R. Corp. and R21 HL086689 to Suzanne Boyce.

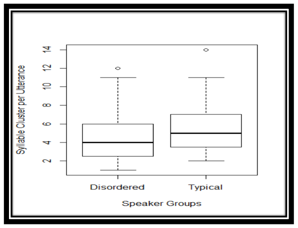
For more information and SpeechMark downloads: www.speechmrk.com

Results

Automated Analysis

- A logistic regression model was used to examine the significance of syllabic clusters in predicting the disorder status.
- The model indicated that syllable cluster per utterance is a significant predictor for the disorder status (Wald $z = 2.619$, $df = 171$, $p = 0.00883$).

Fig. 4 *Box plot of Syllable Cluster per Utterance for typical and disordered speaker groups.*



Automated vs. Conventional Approach

A Spearman’s rank correlation test was conducted to examine the relationship between the automatic measure (i.e. syllable cluster per utterance measure) and hand measure (i.e. Word Complexity Measure). The test indicated that the correlation is moderate, but statistically significant ($r_s = 0.41$, $p < 0.01$).

Discussion

Syllable cluster per utterance was found to be a significant predictor for disorder status and may potentially serve as a diagnostic marker for children at risk for speech-language disorders. The automatic measurement of articulatory complexity moderately agrees with a conventional hand measure of word complexity. A limitation of this pilot study is the small sample size. More speakers are being recruited to further test this automatic approach to tracking change in syllabic complexity with age.

REFERENCES

1. Nathani, S., Ertmer, D. J., & Stark, R. E. (2006). Assessing Vocal Development in Infants and Toddlers. *Clinical Linguistics & Phonetics*, *20*(5), 351–369.
2. Oller, D. K. (2000). *The emergence of the speech capacity*. Psychology Press.
3. Oller, D. K., Eilers, R. E., Neal, A. R., & Schwartz, H. K. (1999). Precursors to speech in infancy: the prediction of speech and language disorders. *Journal of communication disorders*, *32*(4), 223-245.
4. Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children*. *Journal of child language*, *38*(01), 1-34.
5. Flipsen Jr, P. (2006). Measuring the intelligibility of conversational speech in children. *Clinical linguistics & phonetics*, *20*(4), 303-312.
6. Oller, D. K., Niyogi, P., Gray, S., Richards, J. A., Gilkerson, J., Xu, D., ... & Warren, S. F. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences*, *107*(30), 13354-13359.
7. Xu, D., Richards, J. A., & Gilkerson, J. (2014). Automated Analysis of Child Phonetic Production Using Naturalistic Recordings. *Journal of Speech, Language, and Hearing Research*, *57*(5), 1638-1650
8. Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America*, *111*(4), 1872-1891
9. Liu SA. Landmark detection for distinctive feature-based speech recognition. *The Journal of the Acoustical Society of America.* 1996;100(5):3417-3430.
10. Boyce, S., Fell, H. Wilde, L., MacAuslan, J. (2011): *Automated Tools for Identifying Syllabic Landmark Clusters that Reflect Changes in Articulation*”, Proceedings from MAVEBA: 7th Annual Workshop for Models and Analysis of Vocal Emissions for Biomedical Applications
11. Boyce S, Fell HJ, MacAuslan J. SpeechMark: Landmark Detection Tool for Speech Analysis. Paper presented at: INTERSPEECH2012.
12. Stoel-Gammon, C. (2001). Transcribing the speech of young children. *Topics in language disorders*, *21*(4), 12-21
13. Fell, H. J., MacAuslan, J., Ferrier, L. J., & Chenausky, K. (1999). Automatic babble recognition for early detection of speech related disorders. *Behaviour & Information Technology*, *18*(1), 56-63.
14. Fell, H., MacAuslan, J., Ferrier, L. J., & Worst, S., Chenausky, K., (2002). *Vocalization Age as a Clinical Tool*. Proceeding from ICSLP: International Conference on Speech Processing.
15. Stoel-Gammon, C. (2010). The word complexity measure: Description and application to developmental phonology and disorders. *Clinical linguistics & phonetics*, *24*(4-5), 271-28
16. Martin, B., & Carle, E. (1996). *Brown bear, brown bear, what do you see?.* Board book ed. New York: H. Holt.