My Vowels Matter: Formant Automation Tools for Diverse Child Speech

Hannah Valentine[1], *Joel MacAuslan*[2], *Maria Grigos*[1], *Marisha Speights*[3]

[1]New York University, New York, NY, USA
[2]Speech Technology and Applied Research Corporation, Lexington, MA, USA
[3]Northwestern University, Chicago, IL, USA

hannah.valentine@nyu.edu, joelm@staranalyticalservices.com, maria.grigos@nyu.edu, marisha.speights@northwestern.edu

Abstract

Tools to automate formant measurement in vowels have been developed recently, but they have not been tested on pediatric speech samples. Critically, child speech includes unique acoustic challenges including high fundamental frequencies, wide formant bandwidths, more variable formant values, and increased subglottal coupling relative to adult speech. More importantly, these tools have not been tested on the diverse linguistic variations spoken by children. This study compares three tools for automatic formant estimation: Voweltine, Fast Track, and SpeechMark. The tools are tested on vowel productions from a young child with a speech sound disorder from a Black-identifying family. Benefits and tradeoffs of each automation tool are discussed.

Index Terms: acoustic, formant, automation, diversity, child

1. Introduction

Pediatric vowel productions are challenging to research because of their unique articulatory and acoustic considerations. Children have high fundamental frequencies, wide formant bandwidths, highly variable formant values, and increased subglottal coupling relative to adult speech [1]. Existing methods for measuring formants from child speech require manual processing at multiple steps [2, 3]. In the era of big data, these methods are too time-consuming to be functional.

In recent years, tools have been developed to automate the process of vowel formant measurement and tracking. However, these tools have not yet been widely applied to child speech. More critically, these tools haven't been tested on children of color, children with disabilities, and children with intersectional identities. This study demonstrates three automation tools for formant measurement on diverse pediatric linguistic variation.

2. Tool descriptions

2.1. Voweltine

Voweltine[1] (v. 1.0beta) is a Praat®[2] script for automating formant measurements for monophthongs from disordered pediatric speech. The tool runs on a sound file with annotated vowel segments. Voweltine compares linear predictive coding (LPC) parameters to fit straight formant tracks by minimizing aggregated standard deviation. It also identifies and extracts mean F1 and F2 values from the steadiest 30ms of the vowel. Outputs include a text grid indicating steady state and best LPC filter order, mean F1 and F2 values, and analytics on LPC fit.

2.2. Fast Track

Fast Track© is a Praat plug-in for smoothing formant tracks across monophthongs and diphthongs [4]. It uses regression modeling to compare LPC parameters. Fast Track runs on vowel-only sound files. Outputs include F1-F4 measurements every 2ms over the vowel, images of the formant tracks overlaid on the spectrogram, comparison images of other LPC settings, vowel plots, and residual errors for evaluating LPC fit.

2.3. SpeechMark

SpeechMark® (v. 2.0beta) is a MATLAB®[3] toolbox developed by Speech Technology and Applied Research [5]. Using acoustic landmarks [6], it identifies the vowel segments within the sound file. It analyzes the portion of the vowel with the maximum harmonic power. Manual identification of vowels is not required for this tool. SpeechMark uses fixed LPC settings for children but is specifically designed to prevent interference from f0 and subglottal resonances. Outputs include annotated waveforms and power spectra, vowel space area and volume plots [1], F1-F3 values, and a tool confidence score.

3. Methods

3.1. Speaker

The speaker is a girl aged 5;2 living in Ohio. Her family identifies as Black/African American. She was observed using features of African American English and Southern American English. She was also diagnosed with a speech sound disorder by an ASHA-certified speech-language pathologist.

3.2. Stimuli

Recordings come from the Speech Exemplar and Evaluation Database (SEED) [7]. All thirty-five single syllable words from the Clinical Assessment of Articulation and Phonology, 2nd Edition. Auditory inspection of the data revealed vowel errors including substitutions, distortions, resonance errors, and elongations.

---

[1] Valentine, H., Carozzi, G., Grigos, M. (2022) *Automating Vowel Formant Estimations for Disordered Pediatric Speech Samples* [Special session]. Conference on Motor Speech, Charleston, SC.

[2] Boersma, Paul & Weenink, David (2021). Praat: doing phonetics by computer [Software]. Version 6.1.41 `http://www.praat.org/`

[3] The MathWorks, Inc. (2022). MATLAB version: 9.13.0 (R2022b). `https://www.mathworks.com`

4. Demonstration

4.1. Tutorial

Data preparation required for each tool is shown. Then, the three tools are executed on the speaker's vowel productions. Outputs are discussed. Sample outputs from a vowel production with an error ("bed," produced /bijɛ/), are displayed below.

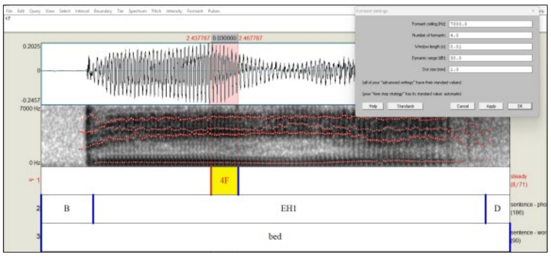

Figure 1: *Example of Voweltine output*

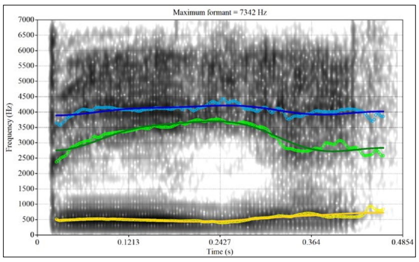

Figure 2: *Example of Fast Track output*

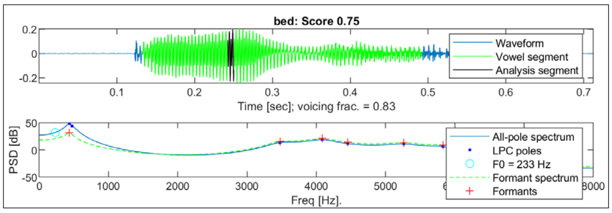

Figure 3: *Example of SpeechMark output*

4.2. Tool comparison

Figure 4 compares tool outputs. Differences between outputs per token were measured by averaging pairwise comparisons of the F1 and F2 log-distance. Blue shows high agreement. Vowels of low F1 and both high and low F2 have the least agreement (red). Tokens colored grey failed for at least one tool.

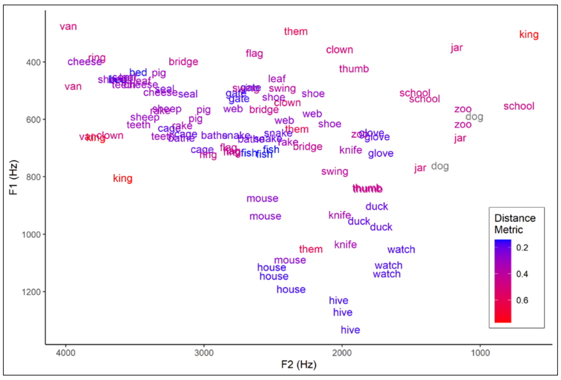

Figure 4: *Comparison of Tool Outputs*

5. Discussion

Several automation tools are available for vowel formant measurement. The tools run on different software platforms, require different data preparation, and have different degrees of user-friendliness. Before selecting a tool, users should consider participant population (e.g., race, age, ability status), vowel type (e.g., monophthongs, diphthongs), data format (e.g., single word recordings, connected speech recordings), research goals (e.g., tracking formants over time, avoiding subglottal resonances), and desired outcomes (e.g., number of formants, static versus dynamic measurements). These tools represent an important step towards future-proofing the field of communication sciences and disorders for the big data era. With these tools, researchers can process large amounts of vowel data in a fraction of the time.

However, it is essential that these tools are properly validated on a wide variety of speakers before implementation. Child speech is acoustically unique. Tools may require further development to process pediatric vowel productions. Additionally, linguistic variation, including dialectical differences and differences in ability status, must be accounted for. This study tested the usability of three tools on pediatric speech with linguistic variation. Results suggest the tools output similar measurements for some tokens and vastly different measurements for other tokens. Future work should validate tool outputs against manual measurement, using vowel productions from a larger corpus of pediatric speech that is representative of real-world speech diversity.

6. Acknowledgements

The authors would like to thank Santiago Barreda, the developer of Fast Track, for consulting on this project. SpeechMark research and development were supported in part by the United States National Institutes of Health (5R44DC010104-04, 3R44DC010104-03S1)

7. References

[1] Vorperian, H. K., & Kent, R. D. (2007). Vowel Acoustic Space Development in Children: A Synthesis of Acoustic and Anatomic Data. *Journal of Speech, Language, and Hearing Research, 50*(6), 1510–1545.

[2] Vallabha, G. K., & Tuller, B. (2002). Systematic errors in the formant analysis of steady-state vowels. *Speech Communication, 38*(1–2), 141–160.

[3] Derdemezis, E., Vorperian, H. K., Kent, R. D., Fourakis, M., Reinicke, E. L., & Bolt, D. M. (2016). Optimizing Vowel Formant Measurements in Four Acoustic Analysis Systems for Diverse Speaker Groups. *American Journal of Speech-Language Pathology*, 25(3), 335–354.

[4] Barreda, S. (2021). Fast Track: fast (nearly) automatic formant-tracking using Praat. *Linguistics Vanguard, 7*(1).

[5] Boyce, S., Fell, H., & MacAuslan, J. (2012). SpeechMark: Landmark detection tool for speech analysis. In *Thirteenth Annual Conference of the International Speech Communication Association.* Available at: `https://www.speechmrk.com`

[6] Stevens, K. N. (1980). Acoustic correlates of some phonetic categories. *The Journal of the Acoustical Society of America, 68*(3), 836-842.

[7] Speights Atkins, M., Bailey, D. J., & Boyce, S. E. (2020). Speech exemplar and evaluation database (SEED) for clinical training in articulatory phonetics and speech science. *Clinical Linguistics & Phonetics, 34*(9), 878-886.