Preliminary steps to validate Audapter, a software application for online tracking of children's formants

Bloorview RESEARCH INSTITUTE

Holland Bloorview Kids Rehabilitation Hospital

Stephanie Cheung MASc[1,2], Kristen Thompson[1], Silvia Orlandi PhD[1],
Yana Yunusova PhD CCC-SLP[3,4,5], Deryk Beal PhD CCC-SLP[1,2,4,5]

Institute of Biomaterials & Biomedical Engineering UNIVERSITY OF TORONTO

[1] Holland Bloorview Kids Rehabilitation Hospital; [2] Institute of Biomaterials and Biomedical Engineering, University of Toronto;
[3] Sunnybrook Health Sciences Centre; [4] Rehabilitation Sciences Institute, University of Toronto; [5] Department of Speech Language Pathology, University of Toronto

Introduction

- Real-time digital tracking of children's formants is difficult to achieve accurately [1].
- Audapter is a commonly-used research software application for real-time tracking and manipulation of adults' [2, 3] and children's [4, 5] formants.
- Audapter's accuracy has not previously been validated, but is important to establish.

Objectives

1. To assess the accuracy and margin of error with which Audapter tracks formants in natural and synthesized samples of children's speech.
2. To determine the speech signal parameters correlated with tracking accuracy.

Methods

Process:

- Speech samples were processed in Audapter.
- Formants were extracted and compared to reference values to determine accuracy.

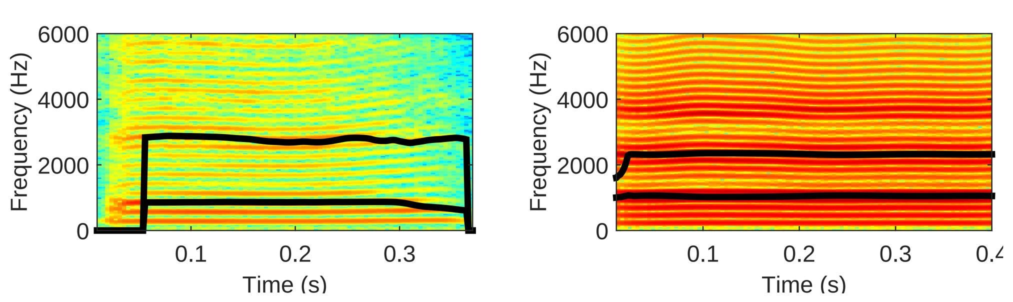

Figure 1: Spectra for (a) natural and (b) synthesized samples with Audapter-detected formants overlaid.

Natural speech samples:

- Obtained from North Texas Vowel Database [6].
- /hɛd/, /hæd/, and /hɪd/ spoken by 3-year-old, 5-year-old, and 7-year-old children.
- Reference values: Database formant values.

Synthesized speech samples:

- Generated with Children's Vowel Synthesizer [7].
- /ɑ/, /æ/, /ʌ/, /ɛ/, /e/, /ɪ/, /i/, /o/, /ʊ/, /u/ for a preschool and school-aged child.
- Reference values: Ground truth formant values.

Analyses:

Table 1: Analyses performed to assess software performance.

Analysis	Nat.	Syn.
Detection rate and magnitude of error	✓	✓
Correlation between detected and reference formants	✓	
Correlation between error and signal features (F_0, duration, and F_2:F_1 ratio)	✓	

Results: Natural samples

- Formants were continuously tracked in 77% of samples, but discontinuously tracked in 6%. Audapter was unable to detect the presence of formants in 17% of samples (Table 2).

Table 2: Detection of the presence of formants in natural speech samples using Audapter

Sample	Continuous	Discontinuous	Undetected	Total
/hɛd/	$N = 65$	$N = 3$	$N = 12$	$N = 80$
/hæd/	$N = 66$	$N = 10$	$N = 16$	$N = 92$
/hɪd/	$N = 50$	$N = 2$	$N = 12$	$N = 64$
All	$N = 181$	$N = 15$	$N = 40$	$N = 236$

- There is no apparent relationship between Audapter's ability to detect the presence of formants and F_0, sample duration, or F_2:F_1 ratio (Figure 2).

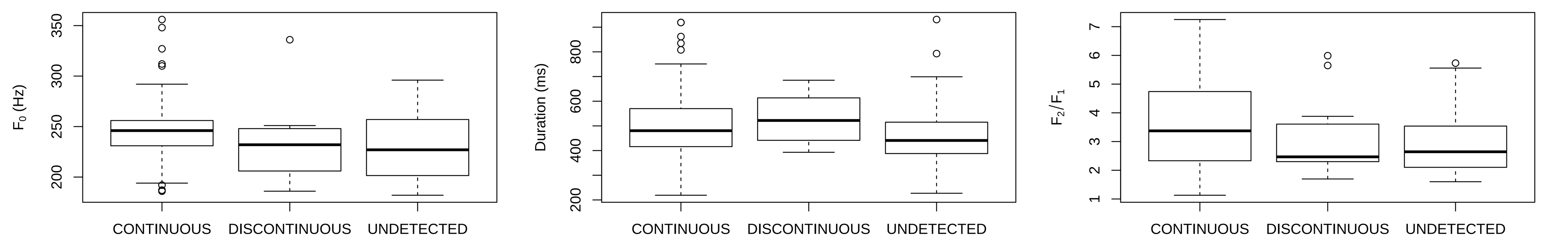

Figure 2: Box plots show the range of (a) F_0, (b) sample duration, (c) F_2:F_1 ratio for continuously-tracked, discontinuously-tracked, and undetected formants .

- Audapter formant measurements were highly correlated with database reference values for F_1 ($r = 0.98, p < 0.01$) and F_2 ($r = 0.96, p < 0.01$). RMS error was 58.03 Hz for F_1 and 183.00 Hz for F_2.
- Only weak correlations were found between F_1 detection error and duration, and F_1 detection error and F_2:F_1 ratio (Table 3).

Table 3: Correlations between Audapter absolute error and database reference values for sample acoustic characteristics. Highlight denotes $p < 0.05$.

	F_0	Duration	F_2:F_1 ratio
$\lvert F_1$ error$\rvert$	$r = -0.02$	$r = 0.15$	$r = -0.21$
$\lvert F_2$ error$\rvert$	$r = -0.03$	$r = 0.04$	$r = 0.08$

Results: Synthetic samples

- Absolute error in Audapter's detection of F_1 and F_2 varied between synthesized vowels (Table 4, Figure 3).
- Formants in the vowel /ɪ/ were not detected, while open-front vowel formants were better tracked.

Table 4: Absolute error in detection of synthesized vowel formants.

	/ɑ/	/æ/	/ʌ/	/ɛ/	/e/	/ɪ/	/i/	/o/	/ʊ/	/u/
$\lvert$Preschooler F_1 error$\rvert$	5	66	73	28	9	69	N/A	48	120	182
$\lvert$Preschooler F_2 error$\rvert$	210	56	51	85	103	2	N/A	12	12	92
$\lvert$School-aged F_1 error$\rvert$	28	6	10	11	1	2	N/A	65	39	68
$\lvert$School-aged F_2 error$\rvert$	35	41	97	91	89	19	N/A	57	46	14

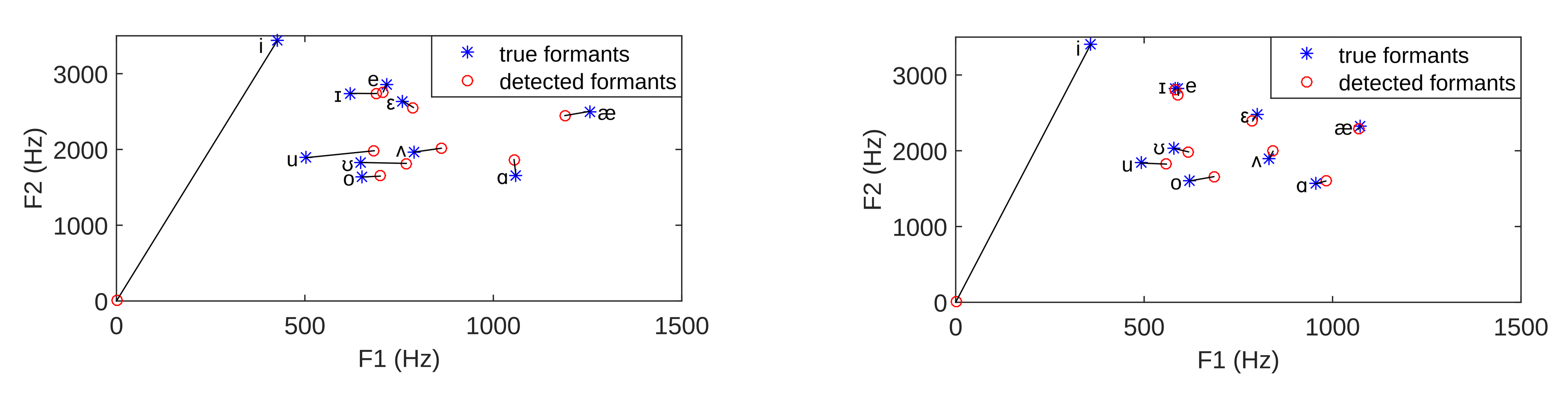

Figure 3: True and detected formant values for (a) preschool and (b) school-aged children's synthesized vowels.

Conclusions

- Some vowels may be more accurately processed (e.g. /æ/) than others (e.g. /ɪ/). This may be a concern when precise tracking is necessary for experimental manipulation of vowel sounds.
- Error margins of 100 Hz for F_1 and 200 Hz for F_2 may be reasonable for most auditory perturbation protocols that shift /hɛd/ to /hæd/.
- Audapter was initialized to search for the exact reference formant values in each sample. For a child whose formants are unknown, error may be higher.
- The relationship between signal features and accuracy is unclear. Other software parameters or unexplored signal features may better explain performance.

Next steps

- Further assessment is needed with additional synthetic samples, as well as natural samples of other vowels.
- The effects of other software parameters and signal features on detection will be explored.
- Audapter's performance should be compared to the performance of offline solutions.

References

[1] Q. Li and M. J. Russell, "Why is automatic recognition of children's speech difficult?," in *INTERSPEECH*, 2001.

[2] S. Cai, D. S. Beal, S. S. Ghosh, M. K. Tiede, F. H. Guenther, and J. S. Perkell, "Weak Responses to Auditory Feedback Perturbation during Articulation in Persons Who Stutter: Evidence for Abnormal Auditory-Motor Transformation," *PLoS ONE*, vol. 7, p. e41830, July 2012.

[3] S. Cai, S. S. Ghosh, F. H. Guenther, and J. S. Perkell, "Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization," *The Journal of the Acoustical Society of America*, vol. 128, pp. 2033–2048, Oct. 2010.

[4] A. Daliri, E. A. Wieland, S. Cai, F. H. Guenther, and S.-E. Chang, "Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter," *Developmental Science*, p. e12521, Feb. 2017.

[5] H. R. Terband and F. J. van Brenk, "Compensatory and adaptive responses to real-time formant shifts in adults and children," 2015.

[6] P. F. Assmann and W. F. Katz, "Time-varying spectral change in the vowels of children and adults," *The Journal of the Acoustical Society of America*, vol. 108, pp. 1856–1866, Sept. 2000.

[7] S. Orlandi and A. Bandini, "Children's Vowel Synthesizer v1.0," 2018.

Acknowledgements

- Dr. Joyce Chen
- Dr. Andrea Bandini
- Keelia Quinn de Launay
- This work is supported by NSERC Discovery Grant RGPIN-2015-05803 and the Ward Family Foundation.