Selecting an acoustic correlate for automated measurement of /ɹ/ production in children

Heather Campbell,[1] Daphna Harel,[2] Elaine Hitchcock,[3] and Tara McAllister[1]

[1] *NYU Steinhardt School of Culture, Education, & Human Development*
[2] *NYU Center for Promotion of Research Involving Innovative Statistical Methodology*
[3] *Montclair Department of Communicative Sciences and Disorders*

Meeting of the Acoustical Society of America, June 26, 2017

Outline

Background

- Speech sound disorders (SSD) can impede academic, social, and psycho-emotional development (Hitchcock et al., 2015).
- For some children, errors are resolve spontaneously, but others require long-term clinical intervention (Flipsen, 2015).
    - May persist through adolescence and, for 1-2% of individuals, into adulthood (Culton, 1986).
- More than 50% of school-based speech-language pathologists (SLPs) report having discharged children with treatment-resistant errors from their caseloads (Ruscello, 1995).

Outline

I: Introduction

1: Why /ɹ/?
2: Visual acoustic biofeedback
3: Automated scoring for /ɹ/
4: Several acoustic measures to consider

II: Methods

1: Data collection
2: Measurement
3: Statistical modeling

III: Results and Discussion

IV: Conclusions and next steps

Why /ɹ/?

- Misarticulation of American English rhotics are the most common and challenging to treat. (Shuster et al., 1995; Ruscello, 1995).
  - Among the latest-acquired speech sounds (Smit et al., 1990).
  - Articulatorily complex: simultaneous anterior and posterior lingual constrictions (Espy-Wilson, 1992) can be achieved with a variety of lingual contours (Delattre and Freeman, 1968).
- Despite articulatory variability, accurate /ɹ/ has stable acoustic properties (Delattre and Freeman, 1968; Hagiwara, 1995)
  - Low third formant frequency (F3) relative to other vowels
  - Second formant frequency (F2) that is close to F3.

Retroflex

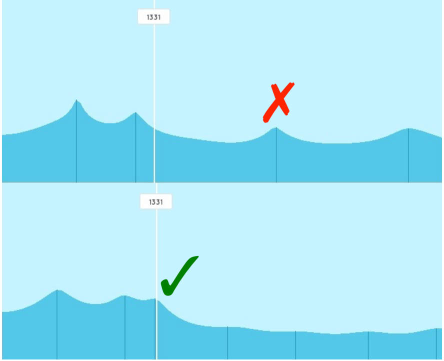
Bunched

Outline

Visual-acoustic biofeedback intervention

- Takes advantage of acoustic consistency of /ɹ/.
  - Display real-time linear predictive coding spectrum representing vocal tract's resonant frequencies.
    - Display target showing correct production of sound.
    - Learner modifies output to align formants with target.
    - Focus is on lowering F3 to match accurate /ɹ/ target.
- Demonstrated efficacy in single case experimental studies.

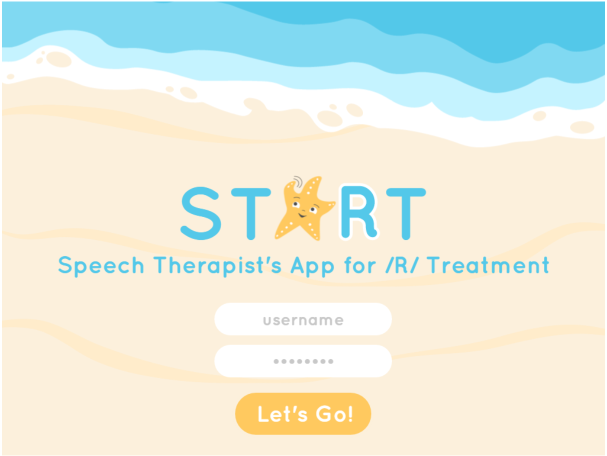
(McAllister Byun, 2017; McAllister Byun and Campbell, 2016; McAllister Byun and Hitchcock, 2012)



Visual-acoustic biofeedback intervention

App-based acoustic biofeedback

- Significant barriers to uptake of tech-based interventions:
  - Cost of the required technology ($2K-$5K).
  - Accessibility and user-friendliness of the technology.
  - Not always a quick solution; may require intensive schedule.
- App under development, in piloting stage (McAllister Byun et al., 2017).



Outline

Motivation for automated scoring

- Gains in treatment do not readily generalize to contexts without biofeedback; longer treatment durations needed (Edeal and Gildersleeve-Neumann, 2011).
- Home practice may help increase the dosage of speech intervention while reducing the strain on SLP resources.
  - **Risk**: Without feedback from SLP, child will counterproductively reinforce incorrect speech patterns.
- **Current need**: Provide valid and reliable automated feedback and track progress during home practice with acoustic biofeedback.

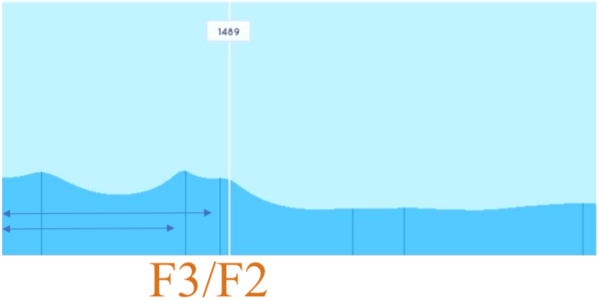
The current study

- **Broad Goal**: Enable home practice with acoustic biofeedback through the incorporation of automated scoring.
- **Which acoustic measure corresponds best with clinician ratings of children's /ɹ/ productions?**
- **Approach**: Compare models that include all possible acoustic values, with and without all possible interactions to find metric that best predicts accuracy.

Outline

Consider raw and derived measures for /ɹ/

- F3: Primary acoustic cue to rhoticity (Espy-Wilson et al., 2000).
  - Low height of F3 differentiates /ɹ/ from acoustically similar sounds such as /l/ and /w/ (Polka and Strange, 1985).
- F2: Secondary acoustic cue to rhoticity (Polka and Strange, 1985).
  - F2 in close proximity to F3.
- Derived within-subject measures reflect the influence of both raw acoustic cues simultaneously (Flipsen et al., 2001; Lee et al., 1999).
  - F3-F2 Distance
  - F3/F2 Ratio



Consider normalization relative to typical speaker data

- Raw and derived measures can be normalized relative to typical speaker data, e.g., Lee et al. (1999) for ages 5-19+.
  - Raw F2 and F3 means and SDs (Lee et al., 1999).
  - Derived F3-F2 and F3/F2 means and SDs (Flipsen et al., 2001).

	Males			Females		
Age	n^b	F3−F2[c]	F3/F2[c]	n^b	F3−F2[c]	F3/F2[c]
5 years	26	797 (343)	1.47 (0.21)	20	643 (210)	1.38 (0.14)
6 years	15	567 (152)	1.37 (0.12)	25	644 (346)	1.35 (0.19)
7 years	19	616 (138)	1.38 (0.09)	32	749 (323)	1.44 (0.22)
8 years	38	517 (175)	1.31 (0.12)	19	669 (497)	1.43 (0.46)
9 years	33	527 (145)	1.34 (0.10)	37	541 (119)	1.31 (0.08)
10 years	40	527 (169)	1.32 (0.11)	24	531 (221)	1.31 (0.13)

Consider interactions with acoustic measures

- Listeners may bring age- and sex-based expectations to a speech rating task that have the potential to interact with the properties of the raw acoustic signal.
  - Perceived age impacts accuracy ratings (Munson et al., 2010).
  - Perceived gender impacts accuracy ratings (Dart, 1991).
- Derivation and normalization may correct for some age- and sex-related differences (Flipsen et al., 2001), but it is unkown whether there are also interactions with these factors.

Outline

Participants

- Children receiving /ɹ/ treatment in 3 biofeedback studies.
  (McAllister Byun et al., 2014; McAllister Byun & Hitchcock, 2012; Hitchcock et al., in press)
  - Normal hearing and oral structure/function.
  - Word probes elicited throughout 8-10 weeks of intervention in a sound-shielded room with the CSL (KayPentax, Model 4150B)

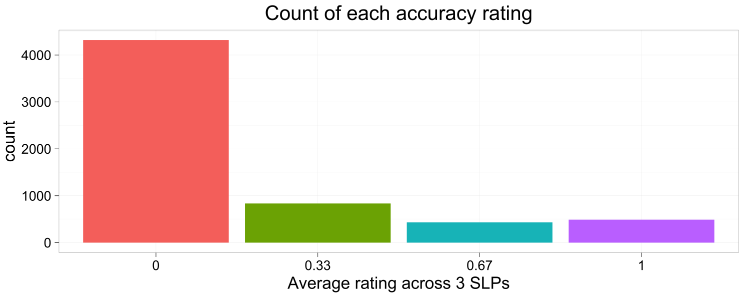
Study	Children	Ages (mean)	Tokens
Acoustic (2012)	11	6-11 (9;0)	2109
Ultrasound (2014)	5	6-9 (7;8)	2926
EPG (2017)	6	6-10 (8;0)	1040
Total	**22**		**6075**

- Varied by phonetic context: Syllabic (808), Post-vocalic (1532), Singleton onset (774), Cluster onset (2961)

Outline

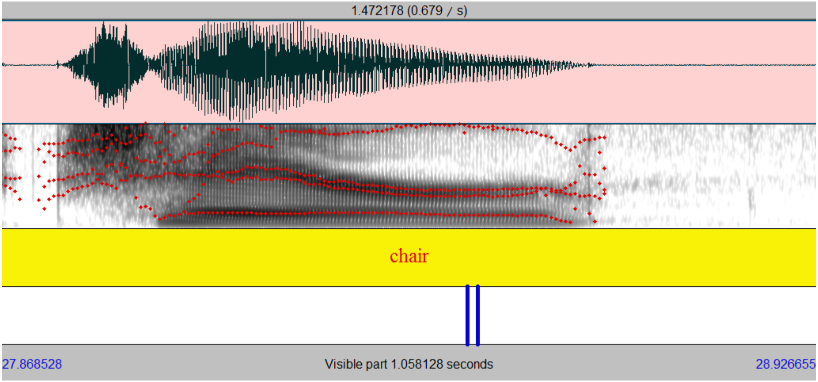
Ratings of perceptual accuracy

- Following the "industry standard" for perceptual rating in speech intervention studies (McAllister Byun et al., 2015), binary ratings were acquired in a blinded randomized fashion from 3 certified SLPs who exhibited at least 80% pairwise agreement.
  - Tokens were rated correct or incorrect.
  - Average across 3 raters was treated as an ordinal scale.
  - Ratings were unequally distributed across accuracy levels.



Acoustic measurement

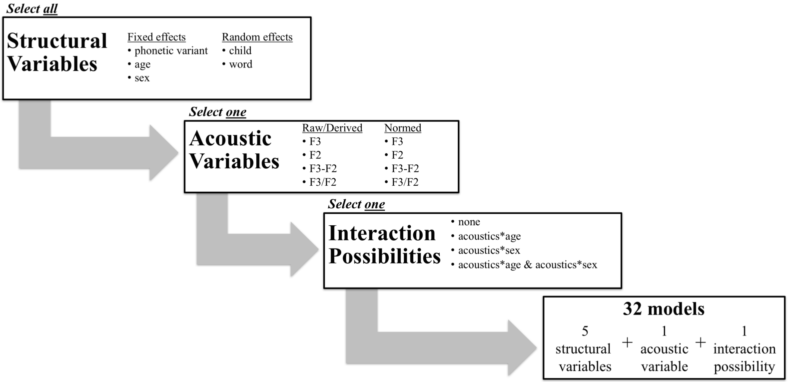
- Trained graduate students measured formant frequencies from the minimum F3 in the rhotic interval of each word using Praat (Boersma and Weenink, 2014).



Outline

Statistical modeling

- A series of ordinal mixed-effects regression models were fit on the aggregated data set while considering the following factors:



Model selection

- Akaike & Bayesian Information Criteria (AIC/BIC) were used to select the best-fitting model.
  - Both take into account the number of predictors (Cohen et al., 2013).
  - BIC penalizes for each predictor, preferring fewer predictors.
  - Select the model with the lowest AIC and BIC.
- All analyses were conducted in R (RStudio, 2016).
  - Data compilation using 'tidyverse' packages (Wickham, 2016).
  - Regression models were fit using the "clmm" function in the 'ordinal' package (Christensen, 2015).

Results

- Controlling for age, sex, and phonetic context, the measure that accounted for the most variance in speech rating was F3-F2 distance normalized relative to a sample of age- and sex-matched speakers.
  - Higher normalized F3-F2 distance was associated with significantly lower accuracy ratings.
  - Best interaction possibility included acoustic variables interacting with both age and sex.
- Cluster onset tokens differed significantly from syllabic and vocalic targets.

Results

- Process for comparing all 32 models.
  - Best models among normalized metrics.

AIC AND BIC FOR ALL 32 MODELS: LOWEST AIC AND BIC FOR EACH INTERACTION
BEST MODELS SHOWN SEPARATELY FOR NORMALIZED AND NON-NORMALIZED METRICS

ACOUSTIC MEASURE INCLUDED IN MODEL	MAIN EFFECTS		MAIN EFFECTS + ACOUSTICS*AGE		MAIN EFFECTS + ACOUSTICS*SEX		MAIN EFFECTS + ACOUSTICS*AGE + ACOUSTICS*SEX	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Normalised F2								
Normalised F3			7900.2	7980.7			7800.2	7887.4
Normalised F3-F2	**7704.9**	**7778.7**	**7680.4**	**7760.9**	**7672.0**	**7752.5**	**7617.3***	**7704.5***
Normalised F3/F2								

Results

- Surprising that normalized version of F3-F2 performed better than non-normalized version of F3-F2.
- Limitations of normative data from Lee et al. (1999):
  - Based on 9-25 individuals in each age/sex group.
  - Speakers from a limited geographic region.
  - Only stressed vocalic /ɝ/ in the word "bird."

TABLE I. Distribution of subjects by age (in years) and gender.

Age	5	6	7	8	9	10	11	12	13	14	15	16	17	18	5–18	25–50
Male	19	11	11	25	23	25	24	22	16	11	11	11	10	10	229	29
Female	13	16	24	11	25	14	19	21	13	10	11	11	9	10	207	27
Total	32	27	35	36	48	39	43	43	29	21	22	22	19	20	436	56

Results

- Process for comparing all 32 models.
  - Best models among normalized metrics.
  - Best models among non-normalized metrics.

AIC AND BIC FOR ALL 32 MODELS: LOWEST AIC AND BIC FOR EACH INTERACTION
BEST MODELS SHOWN SEPARATELY FOR NORMALIZED AND NON-NORMALIZED METRICS

ACOUSTIC MEASURE INCLUDED IN MODEL	MAIN EFFECTS		MAIN EFFECTS + ACOUSTICS*AGE		MAIN EFFECTS + ACOUSTICS*SEX		MAIN EFFECTS + ACOUSTICS*AGE + ACOUSTICS*SEX	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
F2								
F3			7871.1	7951.7			7851.6	7938.9
F3-F2	7752.0	7825.8	*7739.7*	*7820.2*	7753.9	7834.5	7741.7	7828.9
F3/F2								
Normalised F2								
Normalised F3			7900.2	7980.7			7800.2	7887.4
Normalised F3-F2	**7704.9**	**7778.7**	**7680.4**	**7760.9**	**7672.0**	**7752.5**	**7617.3***	**7704.5***
Normalised F3/F2								

Conclusions

- For future automated scoring of children's /r/ productions:
  - If normative data are appropriate, use the externally normalized F3-F2, in interaction with the child's age and sex.
  - Otherwise, we recommend the non-normalized version of F3-F2, in interaction with age only.
- App-based treatment with automated scoring may facilitate increases in treatment dosage by allowing home practice.

Next steps

- Collect more representative normative values, including:
  - A larger sample of children.
  - A more geographically diverse sample.
  - Phonetic contexts other than the syllabic rhotics.
- Improve current aggregated data set:
  - Obtain gradient ratings rather than binary ratings (McAllister Byun et al., 2016; Schellinger et al., 2016; Munson et al., 2012, 2017).
  - Obtain crowd-sourced ratings from naïve listeners (McAllister Byun et al., 2015), which may differ from SLP ratings (Klein et al., 2012).
  - Include potential control for different phonetic context: duration (Klein et al., 2012).

References

Boersma, P. and Weenink, D. (2014). Praat: doing phonetics by computer.

Christensen, R. H. B. (2015). âordinalâ package in r: Regression models for ordinal data via cumulative link (mixed) models.

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge, New York, NY.

Culton, G. L. (1986). Speech disorders among college freshmen: A 13-year survey. *Journal of Speech and Hearing Disorders*, 51(1):3–7.

Dart, S. N. (1991). *Articulatory and acoustic properties of apical and laminal articulations*. PhD thesis, UCLA Working Papers in Phonetics, Los Angeles, CA.

Delattre, P. and Freeman, D. C. (1968). A dialect study of american râs by x-ray motion picture. *Linguistics*, 6(44):29–68.

Edeal, D. M. and Gildersleeve-Neumann, C. E. (2011). The importance of production frequency in therapy for childhood apraxia of speech. *American Journal of Speech-Language Pathology*, 20(2):95–110.

Espy-Wilson, C. Y. (1992). Acoustic measures for linguistic features distinguishing the semivowels /wjrl/ in american english. *Journal of the Acoustical Society of America*, 92(2):736–757.

Espy-Wilson, C. Y., Boyce, S. E., Jackson, M., Narayanan, S., and Alwan, A. (2000). Acoustic modeling of american english /r/. *Journal of the Acoustical Society of America*, 108(1):343–356.

Flipsen, P. (2015). Emergence and prevalence of persistent and residual speech errors. *Seminars in Speech and Language*, 36(4):217–223.

Flipsen, P., Shriberg, L. D., Weismer, G., Karlsson, H. B., and McSweeny, J. L. (2001). Acoustic phenotypes for speech-genetics studies: reference data for residual /ɝ/ distortions. *Clinical Linguistics & Phonetics*, 15(8):603–630.

Hagiwara, R. (1995). *Acoustic realizations of American /r/ as produced by women and men*. PhD thesis, UCLA, Los Angeles, CA.

Hitchcock, E., Harel, D., and McAllister Byun, T. (2015). Social, emotional, and academic impact of residual speech errors in school-aged children: A survey study. *Seminars in speech and language*, 36(4):283–293.

Klein, H. B., Grigos, M. I., McAllister Byun, T., and Davidson, L. (2012). The relationship between inexperienced listeners' perceptions and acoustic correlates of children's /r/ productions. *Clinical linguistics & phonetics*, 26(7):628–645.

Lee, S., Potamianos, A., and Narayanan, S. (1999). Acoustics of childrenâs speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, 105(3):1455–1468.

McAllister Byun, T. (2017). Efficacy of visual-acoustic biofeedback intervention for residual rhotic errors: A single-subject randomization study. *Journal of Speech, Language, and Hearing Research*, 60(5):1175–1193.

McAllister Byun, T. and Campbell, H. (2016). Differential effects of visual-acoustic biofeedback intervention for residual speech errors. *Frontiers in Human Neuroscience*, 10(567):1–17.

McAllister Byun, T., Campbell, H., Carey, H., Liang, W., Park, T. H., and Svirsky, M. (2017). Enhancing intervention for residual rhotic errors via app-delivered biofeedback: A case study. *Journal of Speech, Language, and Hearing Research*, 60(6S):1810–1817.

McAllister Byun, T., Halpin, P. F., and Szeredi, D. (2015). Online crowdsourcing for efficient rating of speech: A validation study. *Journal of communication disorders*, 53:70–83.

McAllister Byun, T., Harel, D., Halpin, P. F., and Szeredi, D. (2016). Deriving gradient measures of child speech from crowdsourced ratings. *Journal of Communication Disorders.*, 64:91–102.

McAllister Byun, T. and Hitchcock, E. R. (2012). Investigating the use of traditional and spectral biofeedback approaches to intervention for /r/ misarticulation. *American Journal of Speech-Language Pathology*, 21(3):207–221.

Munson, B., Edwards, J., Schellinger, S. K., Beckman, M. E., and Meyer, M. K. (2010). Deconstructing phonetic transcription: Covert contrast, perceptual bias, and an extraterrestrial view of vox humana. *Clinical linguistics & phonetics*, 24(4-5):245–260.

Munson, B., Schellinger, S. K., and Edwards, J. (2017). Bias in the perception of phonetic detail in childrenâs speech: A comparison of categorical and continuous rating scales. *Clinical linguistics & phonetics*, 31(1):56–79.

Munson, B., Schellinger, S. K., and Urberg-Carlson, K. (2012). Measuring speech-sound learning using visual analog scaling. *SIG 1 Perspectives on Language Learning and Education*, 19(1):19–30.

Polka, L. and Strange, W. (1985). Perceptual equivalence of acoustic cues that differentiate /r/ and /l/. *Journal of the Acoustical Society of America*, 78(4):1187–1197.

RStudio (2016). Rstudio: integrated development for r.

Ruscello, D. M. (1995). Visual feedback in treatment of residual phonological disorders. *Journal of communication disorders*, 28(4):279–302.

Schellinger, S. K., Munson, B., and Edwards, J. (2016). Gradient perception of childrenâs productions of /s/ and /θ/: A comparative study of rating methods. *Clinical Linguistics & Phonetics*, pages 1–24.

Shuster, L. I., Ruscello, D. M., and Toth, A. R. (1995). The use of visual feedback to elicit correct /r/. *American Journal of Speech-Language Pathology*, 4(2):37–44.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., and Bird, A. (1990). The iowa articulation norms project and its nebraska replication. *Journal of Speech and Hearing Disorders*, 55(4):779–798.

Wickham, H. (2016). 'tidyverse' packages in r: Easily install and load 'tidyverse' packages.

Thank you!

Questions?

heather.campbell@nyu.edu

Acknowledgment

Thanks to members of the Biofeedback Intervention for Speech Lab and the Montclair speech lab for assistance with data collection and processing.