Creaky Voice in Adolescents with Autism Spectrum Disorder

Ethan Weed[1] Riccardo Fusaroli[1] Jessica Mayo[2] Inge-Marie Eigsti[3]

[1]Aarhus University [2]Yale School of Medicine [3] University of Connecticut

Introduction

For over 30 years, people with Autism Spectrum Disorder (ASD) have been described as having not only unusual *prosodic* (e.g."robotic", "flat", "monotone"), but also *vocal* (e.g. "creaky", "harsh", "nasal" and "hoarse") qualities to their speech [1], and there is evidence that neurotypical peers respond negatively to these vocal manerisms [2]. However, there is little consensus on exactly how the prosody and vocal mannerisms of people with ASD differ from typical speech [3]. Although a recent meta-analysis [4] indicated a growing interest in using acoustic measures of speech to quantify these subjective impressions, all studies but one focused on more traditional aspects of prosody, such as pitch and rhythm, and ignored voice quality.

Creaky voice and autism

We focus on creaky voice (vocal fry), which typically occurs when the vocal folds are brought tightly together giving a low, growly quality to the voice [5]. Atypically creaky voice has been reported by several authors as characteristic of the vocal patterns of adults [6], preadolescents [7], and preverbal children [8] with ASD, but has not been investigated acoustically.

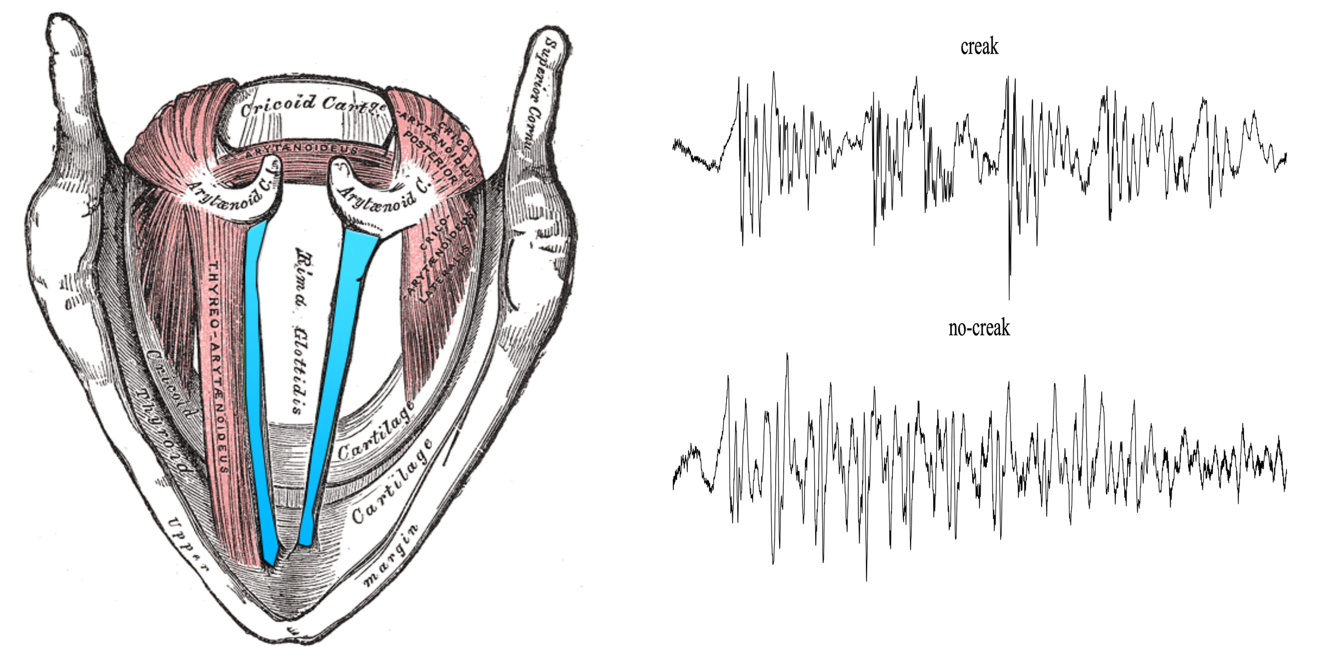

Figure 1: Arytenoid cartilage and vocal folds (shown in blue) and examples of creaky and non-creaky vowels

Table 1: Participant Information

	ASD (15)	TD (15)	p	Cohen's.d/r
Age (years)	14.4 (1.9)	14.0 (1.5)	0.53	0.23
Male: Female	13.02	14.01	0.54	0.11
FSIQ Standard Score (SD)	102 (10)	103 (9)	0.77	0.11
CELF Core Language Standard Score (SD)	109 (10)	116 (10)	0.06	0.70

Methods

Feature Extraction We extracted prosody and voice measures using the Covarep [9] toolbox for Matlab [10] and custom Praat [11] scripts, and recurrence measures of voice creak using the nonlinear Tseries package [12] for R [13]. Automatic detection of voice creak was conducted using the Power Peak Detection method described by Ishi [14] and implemented by Degottex [9]. Pauses were defined as unvoiced periods of syllable length [15] or longer (200 ms.)

Feature Selection We selected common prosodic features (F0, SD of F0, pause duration, and speech rate) measures of voice creak (mean creak, SD of creak, and predictibility of patterns of creakiness (RATIO).

Feature selection was further informed by inspecting intercorrelations, and algorithmic feature selection (t-stats [16])

Model Building We built four Bayesian logistic regression models, using the brms[17] R package. Models were estimated using a Student-t prior distribution, and the included parameters are summarized in Table 2. Model estimation was performed following a bootstrapping procedure in which models were trained on 70% of the data, and generated predictions about the remaining 30%. This procedure was repeated 100 times. Reported accuracy measures are mean accuracy over 100 runs.

Objectives

1. Replicate earlier findings suggesting that acoustic measures of **prosody** can predict ASD diagnosis with a reasonable degree of accuracy
2. Investigate whether adding measures of **voice creak** substantially improves diagnostic classification models built on traditional prosodic features.

Results

Mean accuracy ranged from 66% to 77% (Figure 2). Model 2 had the highest mean accuracy, however we preferred Model 4, as it achieved comparable accuracy with fewer predictors (WAIC, Figure 2). Model accuracy results are shown in Table 3.

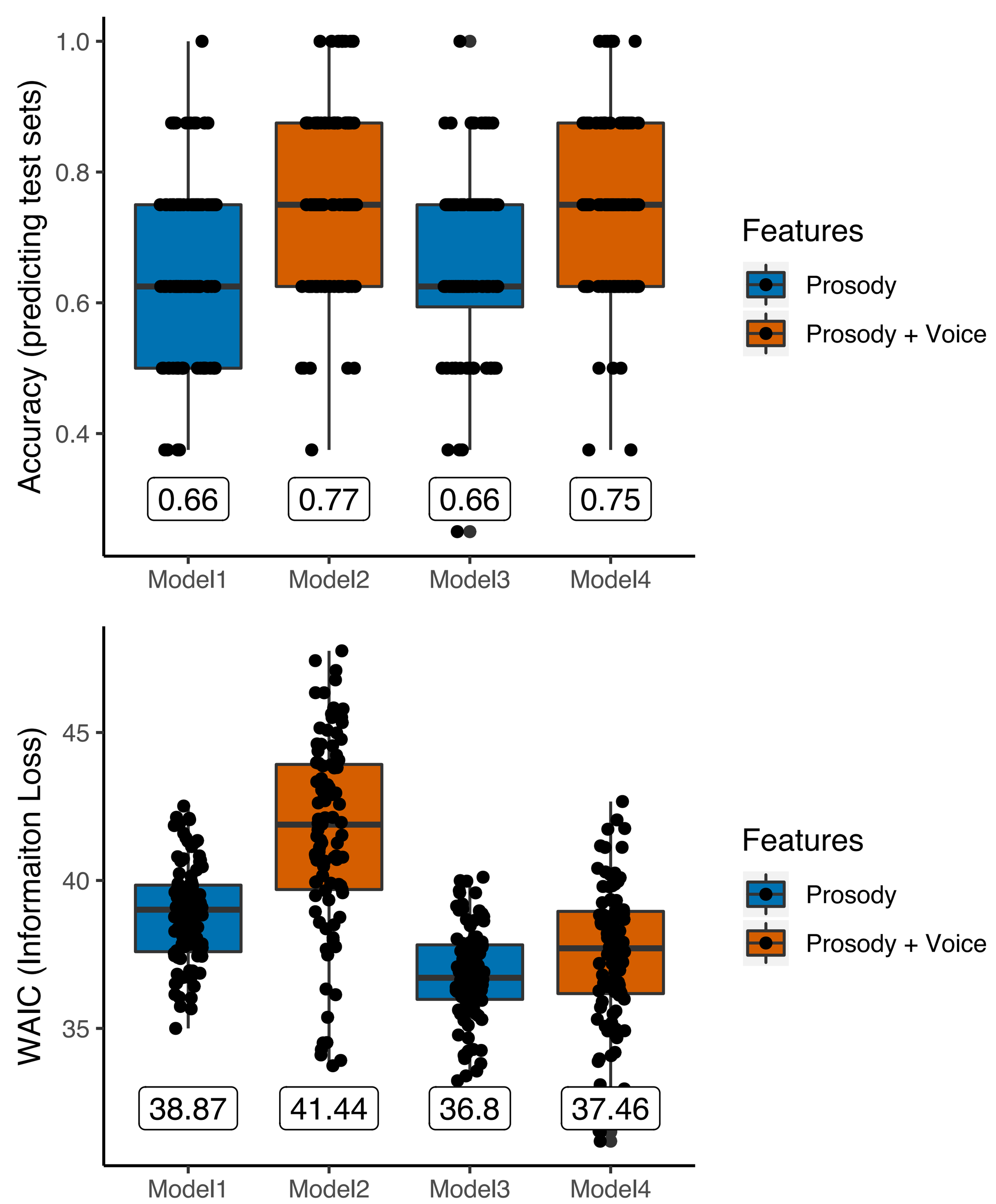

Figure 2: Bootstrapped prediction accuracy and model quality (Watanabe–Akaike Information Criterion, WAIC) with 100 runs

Table 2: Models

Model	Mean F0	F0 SD	Pause Duration	Speech Rate	Mean Creak	Creak SD	Creak RATIO
Mod 1	X	X	X	X			
Mod 2	X	X	X	X	X	X	X
Mod 3		X	X	X			
Mod 4		X	X	X		X	X

Table 3: Model accuracy measures (bootstrapped over 100 runs predicting test sets

	Model 1	Model 2	Model 3	Model 4
Mean accuracy	65.6% (CI: 63-68)	76.6% (CI :74-79.2)	66.1% (CI:63.4-68.8)	74.7% (CI: 72.2-77.2)
Mean sensitivity	63% (CI: 58.5-67.4)	67.5% (CI: 62.5-72.4)	63.2% (CI:58.9-67.5)	70.5% (CI: 65.7-75.2)
Mean specificity	68% (CI: 63.7-72.7)	85.7% (CI: 82-89.4)	69% (CI: 64.3-73.6)	79% (CI:75.3-82.6)

Beta estimates for the preferred model (Model 4) can be seen in Table 4. Participants with ASD tended to have:

- Lower variation in fundamental frequency (F0) (F0_SD)
- Longer pauses (pause_duration)
- Faster speech (speech_rate)
- More variation in voice creak (creak_SD)
- Less regular patterns of voice creak (creak_RATIO)

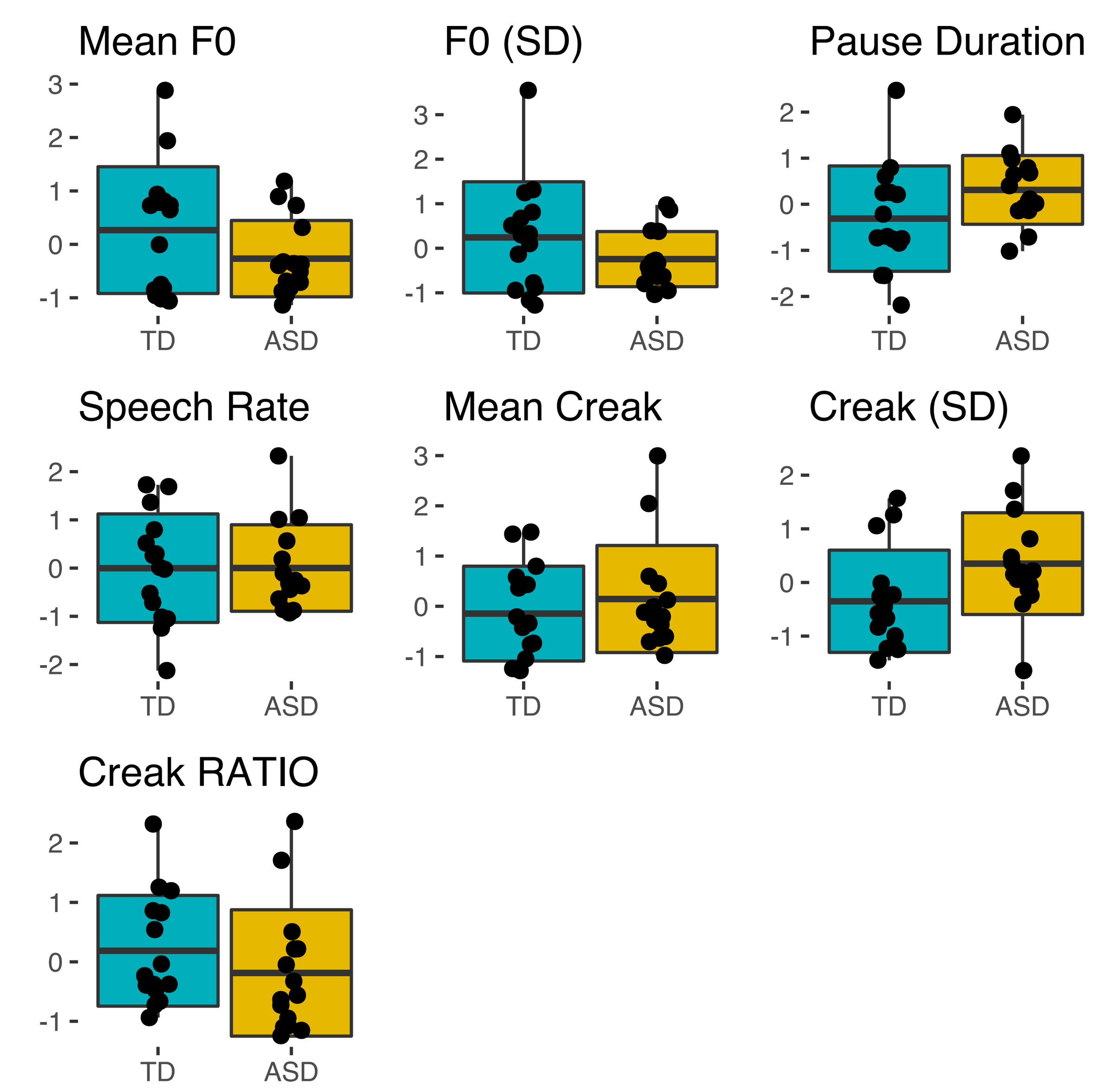

Figure 3: Mean vaules for all predictors (all data)

Table 4: Beta estimates for Model 4 (all data)

Feature	Estimate	Estimated Error	95% CI Lower	95% CI Upper	Effective Sample	Rhat
Intercept	1.66	0.44	0.86	2.60	2164	1
F0_SD	-0.86	0.41	-1.70	-0.11	1998	1
pause_duration	1.30	0.58	0.23	2.52	1968	1
speech_rate	1.48	0.66	0.32	2.89	1596	1
creak_SD	0.85	0.41	0.08	1.70	2242	1
creak_RATIO	-0.41	0.43	-1.28	0.42	2762	1

Conclusions

1. **Objective 1 (replication):** Traditional prosodic measures alone predicted diagnosis with only **66%** accuracy.
2. **Objective 2 (addition of creak)**: Adding measurements of voice quality (in this case, variation and stability of creakiness) improved prediction substantially, bringing accuracy to **75-77%**.

ASD diagnosis could be predicted to some degree using traditional measures of prosody, and prediction was substantially improved by adding measures of voice creak.

Next Steps

1. Replicate these results in other and larger datasets (including other languages and age groups)
2. Connect these results to perceptual impressions of atypical voices
3. Relate these results to measures of symptom severity
4. Investigate the causes of atypical creakiness in ASD.
5. Consider whether atypical creakiness can or shoud be targeted in speech-therapy intervention

References

[1] Eric Schopler and Gary B Mesibov. *Communication problems in autism*. Springer Science & Business Media, 1985.

[2] Noah J. Sasson et al. "Neurotypical Peers are Less Willing to Interact with Those with Autism based on Thin Slice Judgments". In: *Scientific Reports* 7.1 (2017).

[3] J McCann and S Peppé. "Prosody in autism spectrum disorders: a critical review." In: *Int J Lang Commun Disord* 38.4 (2003), pp. 325–350.

[4] R Fusaroli et al. "Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis"." In: *Autism Res* 10.3 (2017), pp. 384–407.

[5] Peter Ladefoged and Sandra Ferrari Disner. *Vowels and consonants*. John Wiley & Sons, 2012.

[6] Sushie Dobbinson, Mick Perkins, and Jill Boucher. "The interactional significance of formulas in autistic language". In: *Clinical Linguistics & Phonetics* 17.4-5 (2003), pp. 299–307.

[7] Mari Wiklund. "Interactional challenges in conversations with autistic preadolescents: The role of prosody and non-verbal communication in other-initiated repairs". In: *Journal of Pragmatics* 94 (2016), pp. 76–97.

[8] Stephen J Sheinkopf et al. "Vocal atypicalities of preverbal autistic children". In: *Journal of Autism and Developmental Disorders* 30.4 (2000), pp. 345–354.

[9] *Covarep: a collaborative voice analysis repository for speech technologies*. Vol. Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, 2014, pp. 960–964.

[10] MATLAB. *version 9.3 (R2017b)*. Natick, Massachusetts: The MathWorks Inc., 2017.

[11] P Boersma and D Weenink. "Praat, a system for doing phonetics by computer". In: (2001).

[12] CA Garcia and G Sawitzki. "nonlinearTseries: Nonlinear Time Series Analysis". In: *R Package Version 0.2* 3 (2015).

[13] R Core Team. *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria, 2013. URL: http://www.R-project.org/.

[14] Carlos Toshinori Ishi et al. "A Method for Automatic Detection of Vocal Fry". In: *IEEE Transactions on Audio, Speech, and Language Processing* 16.1 (2008), pp. 47–56.

[15] Brigitte Zellner. "Pauses and the temporal structure of speech". In: *Zellner, B.(1994). Pauses and the temporal structure of speech, in E. Keller (Ed.) Fundamentals of speech synthesis and speech recognition.(pp. 41-62). Chichester: John Wiley*. John Wiley, 1994, pp. 41–62.

[16] Max Kuhn. "Building predictive models in R using the caret package". In: *Journal of statistical software* 28.5 (2008), pp. 1–26.

[17] Paul-Christian Bürkner. "Advanced Bayesian Multilevel Modeling with the R Package brms". In: *arXiv preprint arXiv:1705.11123* (2017).