Kinematic–Acoustic Coupling as an Objective Measure of Autism Clinical Features

J. P. Momsen[1], A. Naples[1], J. McPartland[1,2]

[1]Yale Child Study Center, Yale University School of Medicine; [2]Center for Brain and Mind Health, Yale University School of Medicine

Background

Speakers naturally move their body in time with the way they speak—gestures are a prime example. Prior work suggests the timing relations between speech and co-speech movement is altered in autism[1-2]

Screening tools for autism do examine behaviors like gestures—focusing on qualities such as how often and naturally speakers use them, or how they relate to what the speaker means. However, observing a speaker in real-time cannot capture more granular details about the way people communicate, such as sub-second timing relations between speech and associated movements of the head, hands, and body.

Here, we ask whether a data-driven approach to characterizing speech-gesture coupling in time can offer an objective supplement to diagnostic practices.

Objectives

The current study evaluates the degree of speech-movement coupling in time within autistic and non-autistic adults.

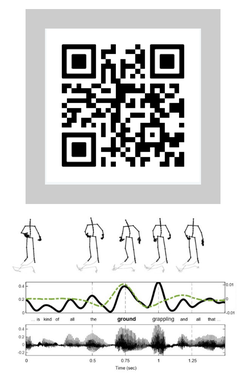
We hypothesize:

Natural co-speech movement patterns made during seated ADOS-2 interviews can account for variance in acoustic properties of their own speech.

Speech-movement coupling strength will negatively correlate with autism features across all participants.

Head movements will be more strongly associated with autism features than other body segments (e.g., the arms).

See for example of naturalistic speech-movement alignment



TLDR

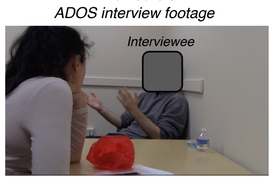
By analyzing video recordings of over 90 clinical screening interviews, we found that the relationship between a speaker's head motion and their speech could predict their ADOS-2 scores.

Across both autistic and non-autistic adults, those who regularly moved their head about 1/5th of a second before moments of vocal stress had fewer autism features on average. Conversely, those with more autism features also had less structured coordination between their head movements and speech over time.

Methods

Video recording taken with consent during 95 clinician-led interviews during the administration of the Autism Diagnostic Observation Schedule (ADOS-2) by a licensed psychologist.

> 70 hours of ADOS interview footage



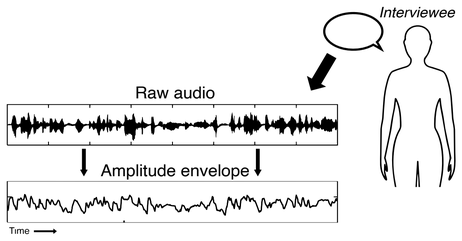
Sample Characterization

Measure Mean(SD)	Non-autistic (n=76)	Autistic (n=19)	Hedge's g
ADOS CS Score	2.73(2.35)	8.00(1.63)	2.37**
Age (years)	25.34(5.52)	25.83(5.96)	0.087
Benton Total	46.35(3.65)	42.42(4.1)	1.049**
RMET Total	26.65(4.93)	22.61(5.48)	0.80*
WASI-II FSIQ	106.49(17.03)	102.74(18.24)	0.217
SRS-2 Raw	49.46(31.03)	77.37(29.90)	0.91**
AQ Total	19.01(9.9)	25.37(7.33)	0.67

★ *p <.05*
★★ *p <.01*
★★★ *p <.001*

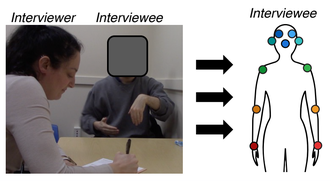
1 Extract interviewee speech

Interview dialogue was diarized, extracted, and decomposed into a low-pass amplitude envelope (10Hz).



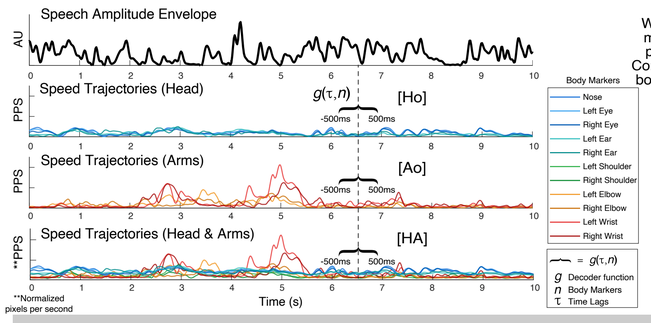
2 Extract interviewee motion during conversation

An open source computer vision program[3] provided 11 2-D position estimates associated with body parts above the waist, which were then transformed into speed trajectories.

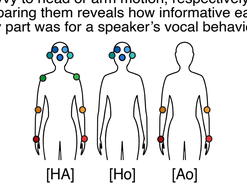


3 Train decoders to predict speech from movement

We computed 3 sets of decoders for each interviewee. The only difference between them was the set of body parts they could use to try and reconstruct the speaker's voice from.



While [HA] decoders were given all 11 body markers, [Ho] and [Ao] decoders were only privvy to head or arm motion, respectively. Comparing them reveals how informative each body part was for a speaker's vocal behavior.

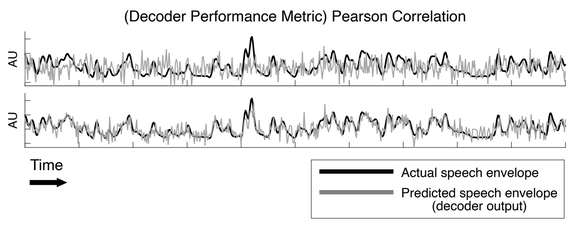


All markers [HA]
Head-only [Ho]
Arms-only [Ao]

A leave-one-out training procedure was applied to create sets of decoder models for each subject, which used ridge regression[4] to predict unit changes in the amplitude of the speaker's voice using only the speed of different body parts that occured within the same second of the interview footage.

4 Evaluate speech-motion coupling strength within interviewees

Pearson correlations between predicted and actual speech envelopes indexed the degree of speech-movement coupling within individuals.



*Example Low Decoder Score
(indicates weaker speech-motion coupling)

*Example High Decoder Score
(indicates stronger speech-motion coupling)

5 Test relationships between coupling strength and symptom measures

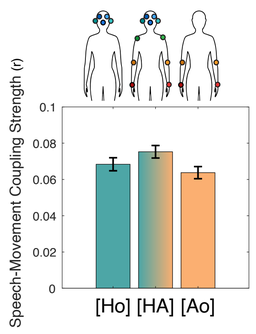
Linear regression models predicting ADOS-2 symptom severity from speech-movement alignment (mean decoder score).

Symptom Measure (ADOS-2) ~ 1 + **Mean Decoder Score** + Age + IQ + Sex + Training Data Quality and Features

Covariates

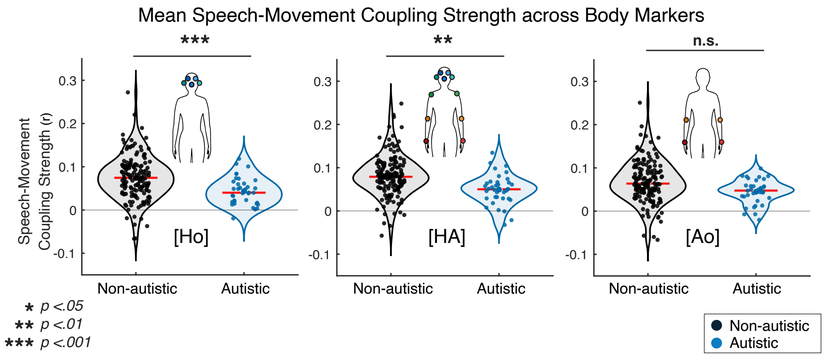
Results

Speech-movement coupling was reduced in autistic compared to non-autistic adults



Average decoder performance for each body part was similar across the entire sample.

Mean Speech-Movement Coupling Strength across Body Markers

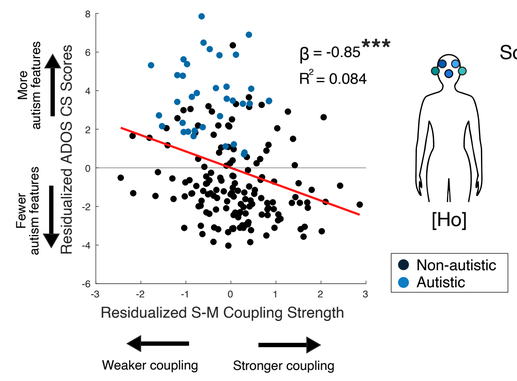


★ *p <.05*
★★ *p <.01*
★★★ *p <.001*

Non-parametric permutation tests (n=5000) revealed that non-autistic participants showed higher decoding performance than autistic participants when trained on all body marker data (Hedges' g = 0.45).
This group difference was more pronounced for the head-only model (Hedges' g = 0.67). In contrast, decoding performance derived from arm motion alone did not differ significantly between groups.

Higher speech-motion coupling was associated with fewer autism symptoms

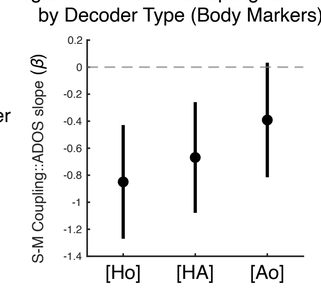
ADOS CS Scores by Speech-Movement Coupling Strength



A linear mixed-effects model using speech-movement coupling strength to predict variance in ADOS-2 Calibrated Severity Scores revealed a significant association, such that higher decoder performance was associated with lower autism symptom scores across all participants (β=-0.67; SE=0.21; p<.001).

The relationship between speech–movement coupling and autism symptom severity was strongest when coupling was quantified using head motion alone (β=-0.85; SE=0.21; p<.0001), and weakest when coupling was derived from arm motion alone (β=-0.39; SE=0.22; p>.05).

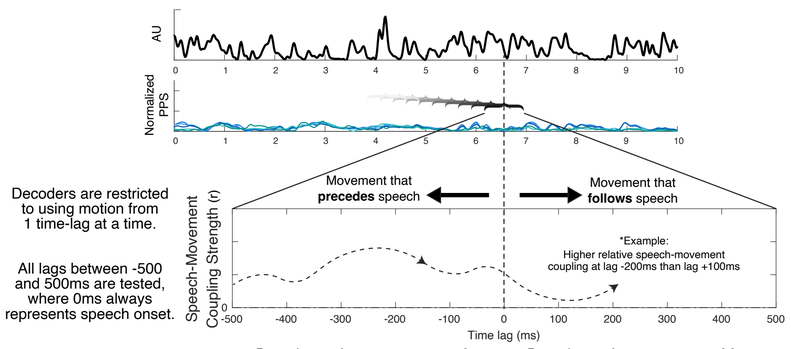
Magnitude of ADOS-Coupling Association by Decoder Type (Body Markers)



Speech-**head movement** coupling held **strongest relationship with clinical measures**

Single-lag Analysis

Single-lag analyses reveal *when* body motion is most informative about speech.



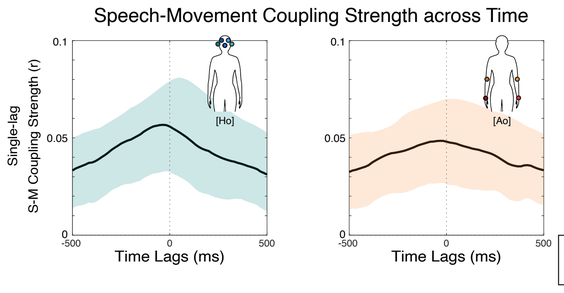
Decoders are restricted to using motion from 1 time-lag at a time.

All lags between -500 and 500ms are tested, where 0ms always represents speech onset.

Decoder performance at **negative lags** reflects **motion occuring prior to upcoming speech**.

Decoder performance at **positive lags** reflects **motion that follows speech**

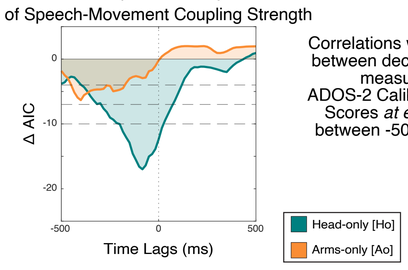
Speech-Movement Coupling Strength across Time



Speech-movement coupling showed a modest peak near t=0ms. This was more prominent for head-only models

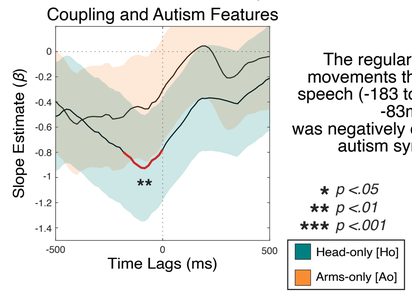
Individuals with stronger coupling between **anticipatory head motion** and vocal output had **fewer autism symptoms.**

Model Fit Improvement by Addition of Speech-Movement Coupling Strength



Correlations were assessed between decoder accuracy measures and ADOS-2 Calibrated Severity Scores *at each time lag* between -500 and 500ms.

Time-Resolved Association Between Speech–Motion Coupling and Autism Features



The regularity of head movements that preceded speech (-183 to 0ms; peak = -83ms) was negatively correlated with autism symptoms.

★ *p <.05*
★★ *p <.01*
★★★ *p <.001*

Conclusions

Our findings demonstrate **weaker speech-movement coupling as a function of increasing autism features.** This relationship held true **even across individuals below the diagnostic threshold for autism**.

This effect was primarily driven by the **relationship** between a speaker's **head movements** and their **speech**[5].

The highest information value for clinical features pertained to head motion that led speech in time by approximately a fifth of a second: the predictability of head movements during this time held a strong inverse relationship with autism symptoms irrespective of age, IQ, self-reported autism traits, and even movement variability (pure motor output).

Differences in **autism** may be more strongly related to continuous coordination of speech with **rhythmically structured head movements** than to canonical manual gestures.

Implications

As the approach debuted here operates on passive conversation recordings, it holds promise to:

*Clarify the nature of nonverbal communication skills in autism

*Supplement **clinical evaluation and longitudinal monitoring** of communicative behavior by using an **objective, data-driven approach.**

Future Directions

Future work may find even more robust results with higher fidelity recordings of the speaker's movements, e.g., through wearable IMUs.

References

[1]Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. Speech communication, 57, 209-232.

[2] de Marchena, A., & Eigsti, I. M. (2010). Conversational gestures in autism spectrum disorders: Asynchrony but not decreased frequency. Autism research, 3(6), 311-322.

[3] TensorFlow. MoveNet: Ultra fast and accurate pose detection model. https://www.tensorflow.org/hub/tutorials/movenet

[4] Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. Frontiers in human neuroscience, 10, 604.

[5] Momsen, J. P., & Coulson, S. (in press). The sensorimotor account of multimodal prosody. In L. Meyer & A. Strauss (Eds.), Rhythms of speech and language. Cambridge University Press.

McPartland Lab

mcp-lab.org
mcp.lab@yale.edu
mcpartland.lab on Instagram