UCSF
University of California
San Francisco

Objective and Subjective Assessment of Facial and Vocal Affect Production in Autistic and Neurotypical Children and Adolescents

Carly Demopoulos[1,2], Linnea Lampinen[1], Cristian Preciado[1], Hardik Kothare[3], & Vikram Ramanarayanan[3]
(1) UCSF Department of Psychiatry and Behavioral Sciences, (2) UCSF Department of Radiology & Biomedical Imaging, (3) Modality.AI

Background

Impairments in nonverbal communication are a defining feature of Autism Spectrum Disorder (ASD).

These impairments can manifest as difficulty with, or complete lack of, communication of emotional states via production of

- **facial affect:** facial expression of emotion
- **vocal affect:** the acoustic cues of vocalization that communicate emotion (not what is said but how it is said)

No standardized objective clinical tools are available to assess expression of affect, and as such, evaluation of this symptom area relies entirely on clinician and/or caregiver subjective judgment.

Objectives

The purpose of this study was to use a novel, standardized Affect Production Task to

- evaluate affect production ability in individuals with autism relative to typically developing control (TDC) participants
- assess agreement between automated objective metrics of facial and vocal expression and human subjective judgment

Methods

Participants

71 children and adolescents ages 8-16 years

- ASD (N=46)
- TDC (N=25)

Affect Production Task (APT)

- quantifies objective facial and vocal affect production ability using audiovisual capture via a virtual dialogue agent (Figure 1)
- elicits a specific, prompted affective facial and vocal response
- does not assess spontaneous emotional response
- isolates a person's ability to intentionally communicate emotions by specifying the emotion to be communicated for each item (happy, sad, angry, or afraid).

APT Subtests

Monosyllabic affective utterances (/oʊ/)

- Noncontextual condition: "Use your face and voice to say 'oh' in a way that seems [happy, sad, angry or afraid]"
- Contextual condition: The participant is read aloud a brief, illustrated, emotional narrative and is asked to say "oh" in a way that conveys the explicitly stated emotion of the character in the narrative (Figure 2).

Noncontextual sentence length affective utterance: "Use your face and voice to say 'I'll be right back' in a way that seems [happy, sad, angry or afraid]"

Procedures

The APT was administered in the laboratory under supervision to ensure effort and task compliance.

Objective metrics of facial and vocal response are captured automatically in all conditions.

- Facial metrics: movement and position of lips, eye opening, eyebrows and mouth (Figure 3)
- Vocal metrics: fundamental and formant frequencies, cepstral peak prominence, timing, pauses, harmonics- and signal-to-noise ratios, intensity, jitter, and shimmer

Subjective Ratings.

- Responses were rated by two research assistants who were blinded to task prompt.
- Raters classified the affective facial and vocal expressions as happy, sad, angry, afraid, or neutral.

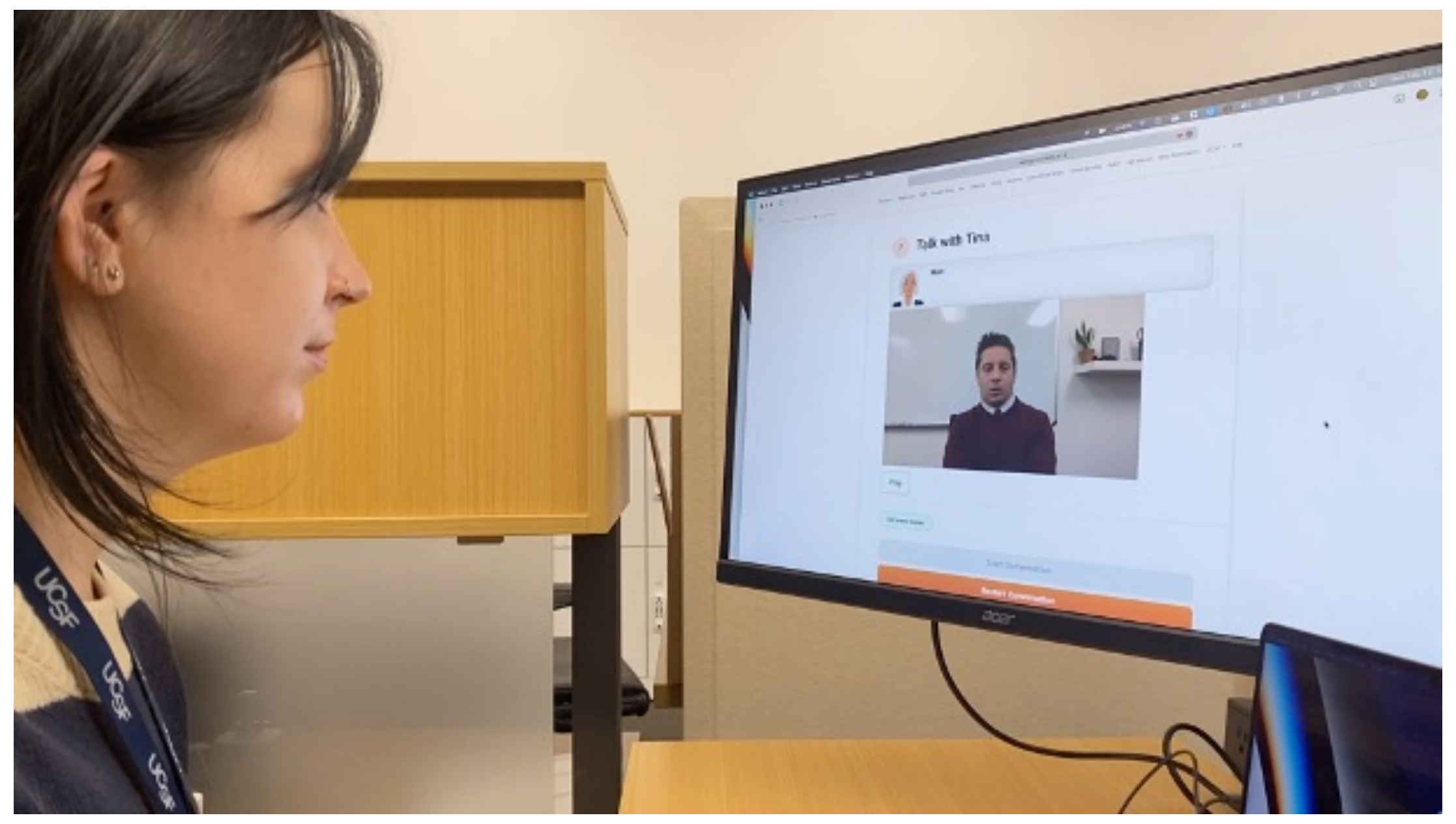

Figure 1 (left). A member of the study team demonstrates completing the APT subtest by interacting with the virtual dialogue system.

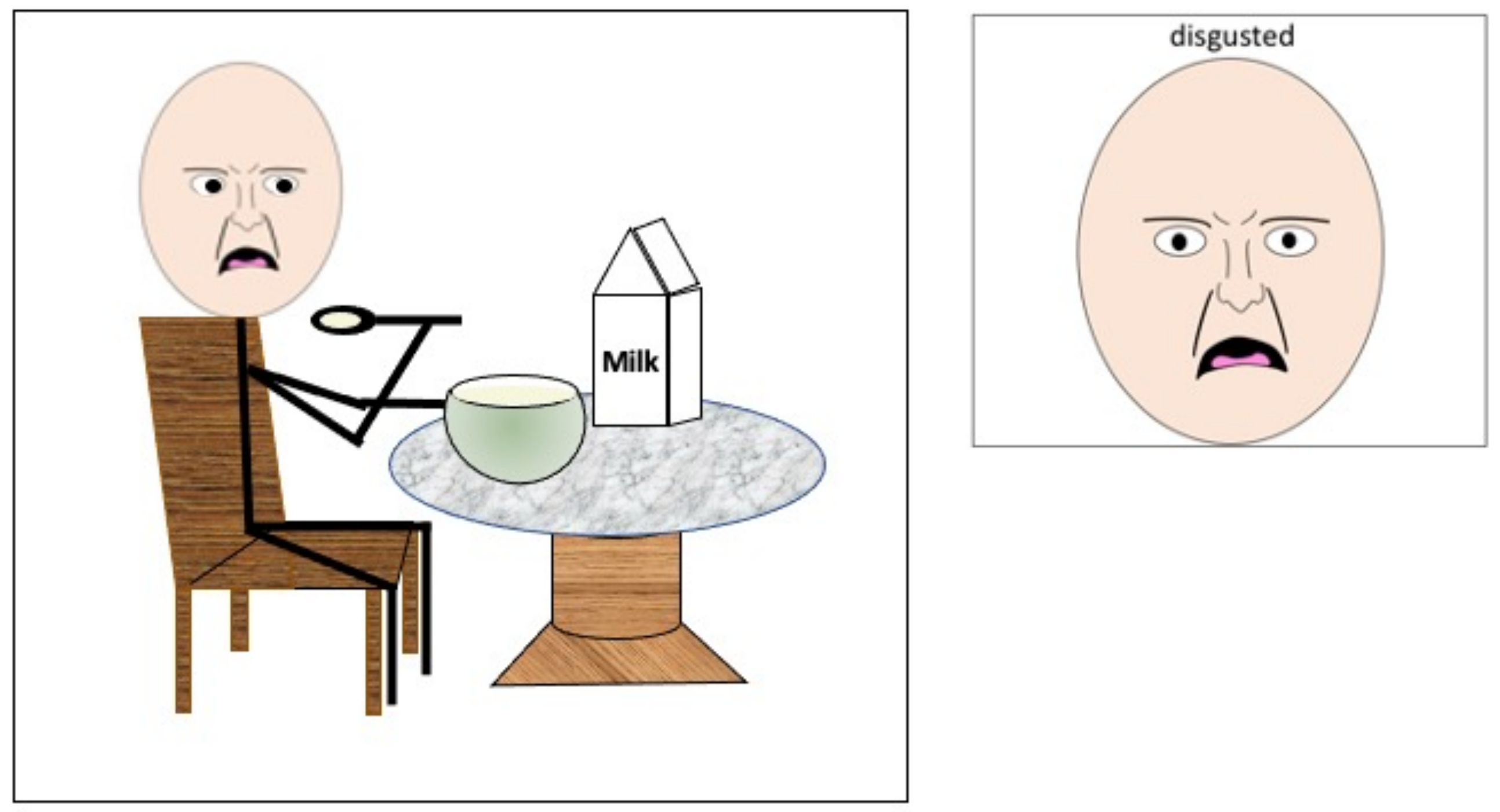

Figure 2 (above) Illustration accompanying an emotional narrative for the teaching trial of the contextual monosyllabic task condition. The participant is read a narrative about a person tasting spoiled milk and feeling disgusted while this image is presented. They are then prompted to use their face and voice to say "oh" in a way that seems disgusted.

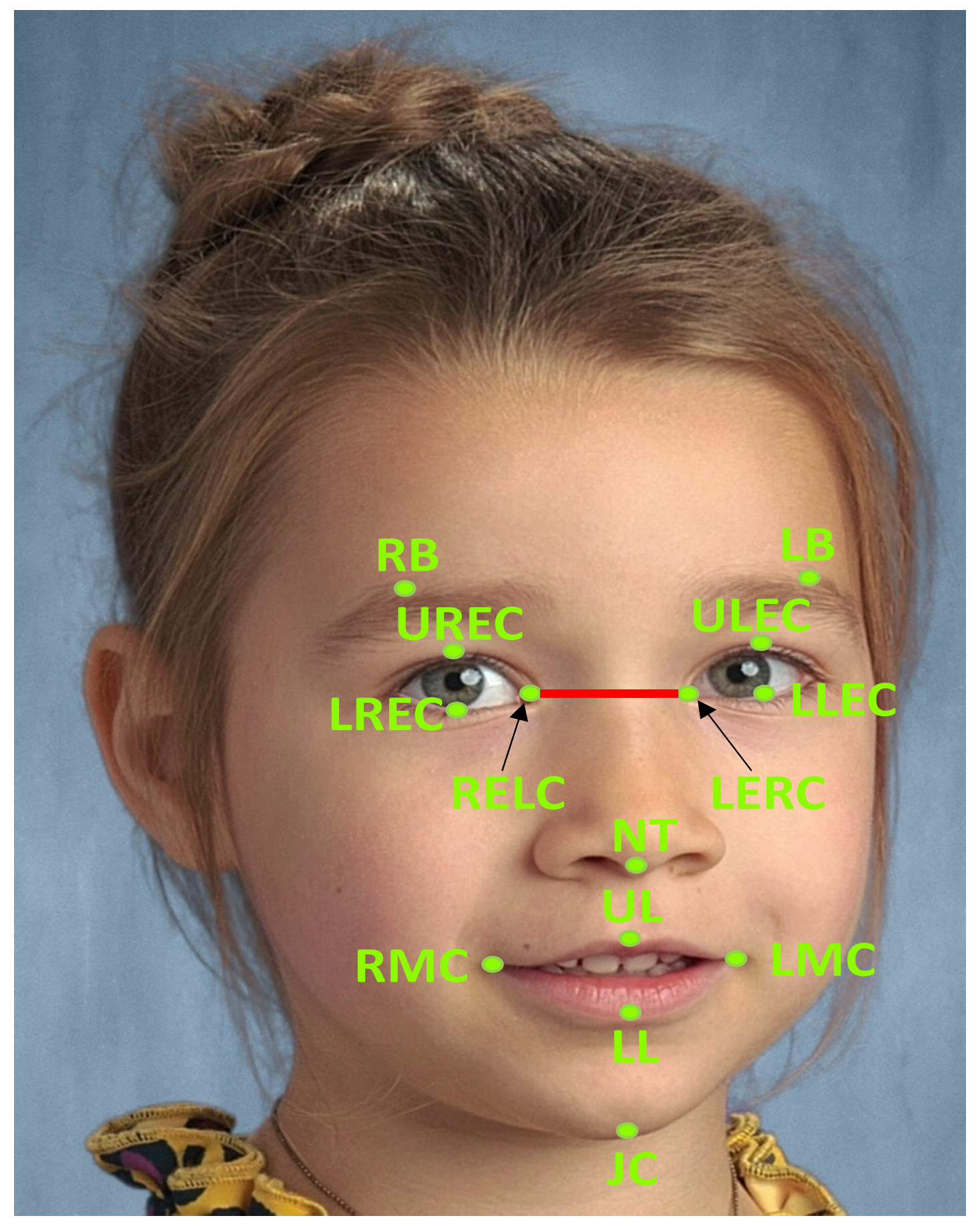

Figure 3. The 14 facial landmarks.
RB: right brow
LB: left brow
UREC: upper right eye center
ULEC: upper left eye center
LREC: lower right eye center
LLEC: left eye center
NT: nose tip
UL: upper lip
LL: lower lip
RMC: right mouth center
LMC: left mouth corner
JC: jaw center

Results

- Nonparametric Kruskal-Wallis tests were performed to examine group differences in % accuracy of the human rater's classifications.
  - ➢ Significant differences in overall accuracy for facial affect production: χ2 (1)=9.760, p=.002
    - ▪ 47% rater accuracy for ASD
    - ▪ 65% rater accuracy for TDC
  - ➢ Significant differences were not identified for overall accuracy in vocal affect production.
    - ▪ 53% rater accuracy for ASD
    - ▪ 63% rater accuracy for TDC
- Stepwise linear regression analyses were performed to assess the prediction of human rater accuracy from objective automated metrics. All regression results were statistically significant:
  - ➢ Facial metrics predicted 60% of the variance in rater accuracy for both noncontextual production tasks (monosyllabic and sentence-length utterances), and 32% of the variance in the contextual monosyllabic production task.
  - ➢ Vocal metrics predicted 41% of the variance in rater accuracy for the noncontextual monosyllabic production task, and 58% of the variance for the noncontextual sentence-length task and the contextual monosyllabic task.

Conclusions

- The automated metrics predicted the accuracy of human raters in classifying facial and vocal affect.
  - ➢ Machine learning approaches may be a viable option for automating and standardizing the quantification of affect production abilities on the APT.
- Facial, but not vocal affect of autistic participants was more difficult for human raters to interpret.
  - ➢ This may be associated with the tendency of many individuals with autism to not look others in the face, resulting in less expertise in communicating via facial affect, whereas voices can be heard regardless of eye contact or visual attention.
  - ➢ This may result in greater opportunity to develop expertise in vocal rather than facial affective communication for the ASD group.

CD was supported by Autism Speaks CAPD Pilot Grant (11637), NIDCD (K23DC016637-01A1; R01DC019167-01A1), and UCSF Weill Institute for Neurosciences (Weill Award for Clinical Neuroscience Research)