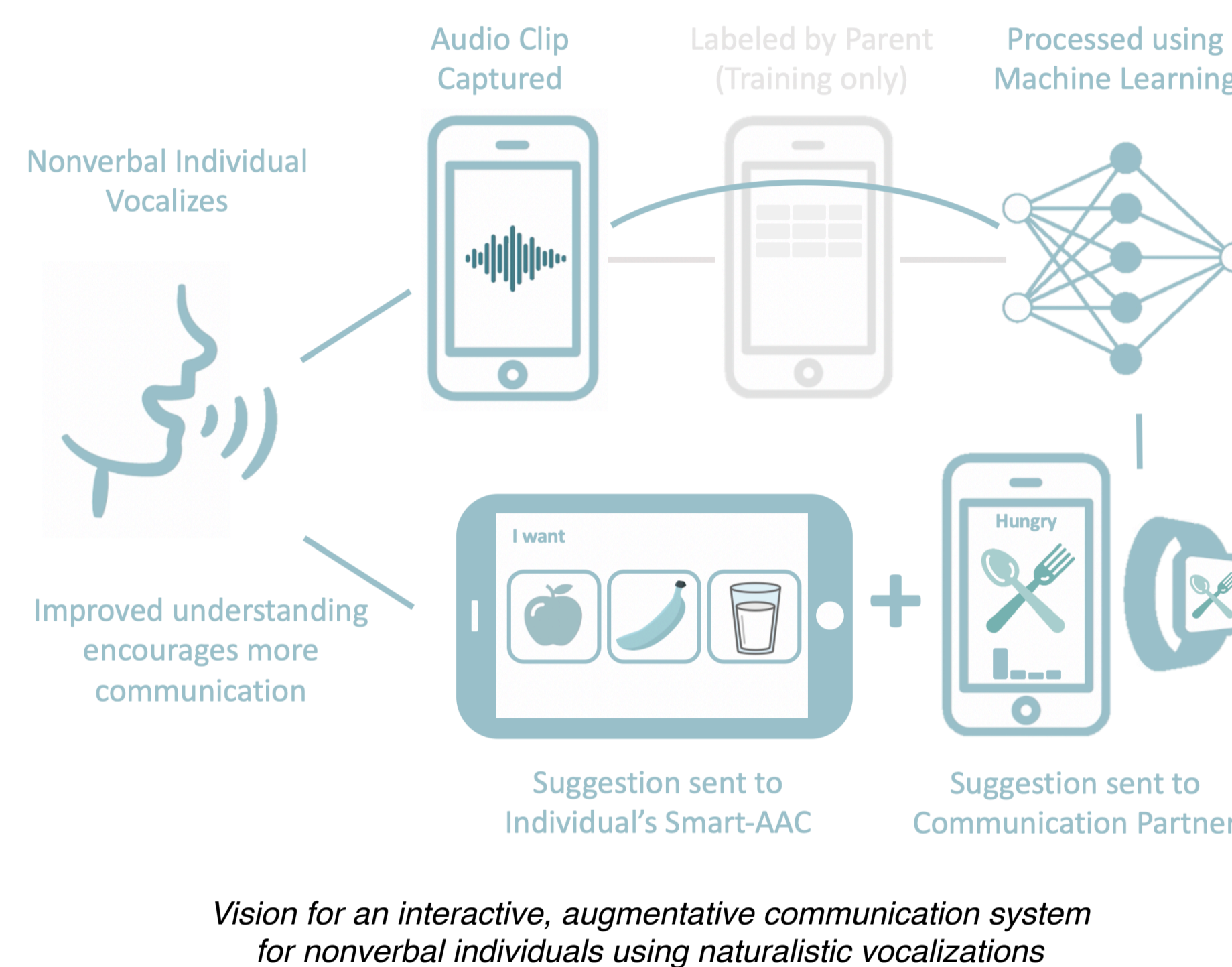# Nonverbal Vocalizations as Speech:
Characterizing Natural-Environment Audio from Nonverbal Individuals with Autism

Jaya Narain* & Kristina Johnson*, Amanda O' Brien, Peter Wofford, Pattie Maes, Rosalind Picard   (*Joint first-authors)

Massachusetts Institute of Technology, {jnarain, ktj, amobrien, peterwof} @mit.edu, {pattie, picard} @media.mit.edu
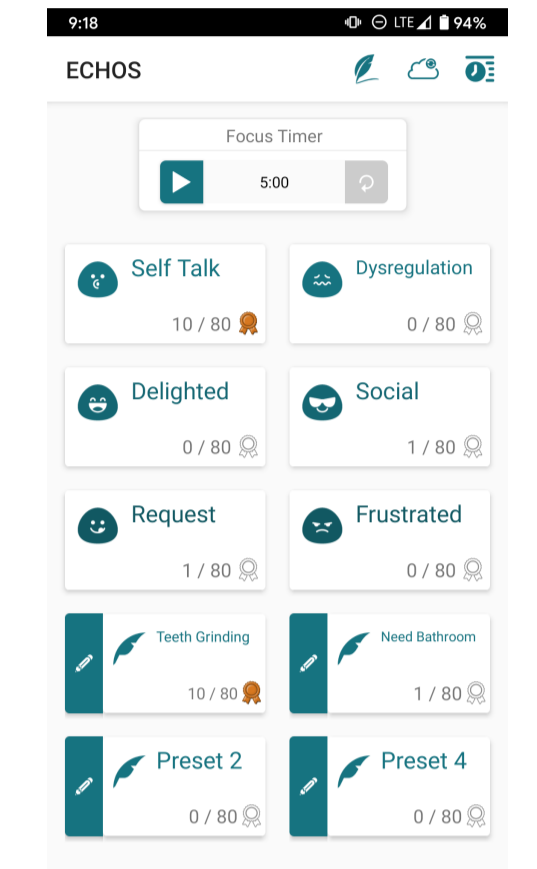
Background

- In the United States alone, **over 1 million people are non- or minimally verbal** (nv/mv), meaning they use zero or fewer than 20 words or word approximations

- Through **non-speech vocalizations**, nv/mv individuals express rich affective and communicative information for a range of functions (e.g., protesting, requesting, affection)

- These vocalizations are **highly idiosyncratic** to the speaker, making interpretation of these utterances challenging for unfamiliar listeners.

- This work is the **first systematic study** on the organic production, meaning, and usage of these sounds

Project Vision

Audio Clip Captured

Labeled by Parent (Training only)

Processed using Machine Learning

Nonverbal Individual Vocalizes

Improved understanding encourages more communication

I want

Hungry

Suggestion sent to Individual's Smart-AAC

Suggestion sent to Communication Partner

Vision for an interactive, augmentative communication system for nonverbal individuals using naturalistic vocalizations

Objectives and Methods

1. **Categorization of nonverbal vocalizations through interviews**

- Conducted interviews with families of nv/mv individuals regarding communication generally and vocal communication, specifically

- The nv/mv individuals were all male and had autism spectrum disorder (ASD)

- Ages: 8 (P0), 19 (P1), and 23 (P2) years old

- P0 also has a rare genetic disorder; P1 also has Down Syndrome

Custom app for "live labeling" by parents

2. **Acquisition and analysis of natural-environment audio recordings with real-time labels.**

- Developed protocol using wearable microphone to comfortably record audio in natural environments

- Created custom app for primary caregivers to label sounds in real time


Parent using labeling app at home

Results

Vocalization Category	Vocalization Interpretation	P0	P1	P2
Self-talk	Self-talk	X	X	X
Dysregulation	Dysregulation	X		X
Frustration	Frustration	X	X	X
	Impatient		X	
	Protest	X		
	Upset		X	
Delight	Glee	X		X
	Happy		X	
	Affection		X	
	Excited		X	
Request	General request	X		X
	Request for a person		X	X
	Request for a high five			X
	Request for the bathroom		X	
Social exchange	Social (general)	X		X
	"Fun participation"		X	
	Greeting			X
Other	Teeth grinding	X	X	
	Singing		X	
	Laughing	X	X	

Family-reported uses of nonverbal vocalizations. The reported presence of a given vocalization type in an individual's non-speech repertoire is indicated by an "X."

- Families reported a **range of nonverbal vocalizations** including

  - **Word approximations** used for more general purposes (e.g., "bɔɪ" for a social interaction)

  - **Consonant-vowel syllables** (e.g., "ma" used for multiple functions)

  - Vocalizations **without phonetic cues** (e.g., laughter)

  - **Sounds** that were often associated with an **affective state** (e.g., teeth grinding related to dysregulation).

- **Distinct vocalization characteristics** are observable

  - Distinctions relate to individual's **emotional or physical state** and **communicative intent**

  - Variations related to **pitch**, **duration** and other **acoustic features**
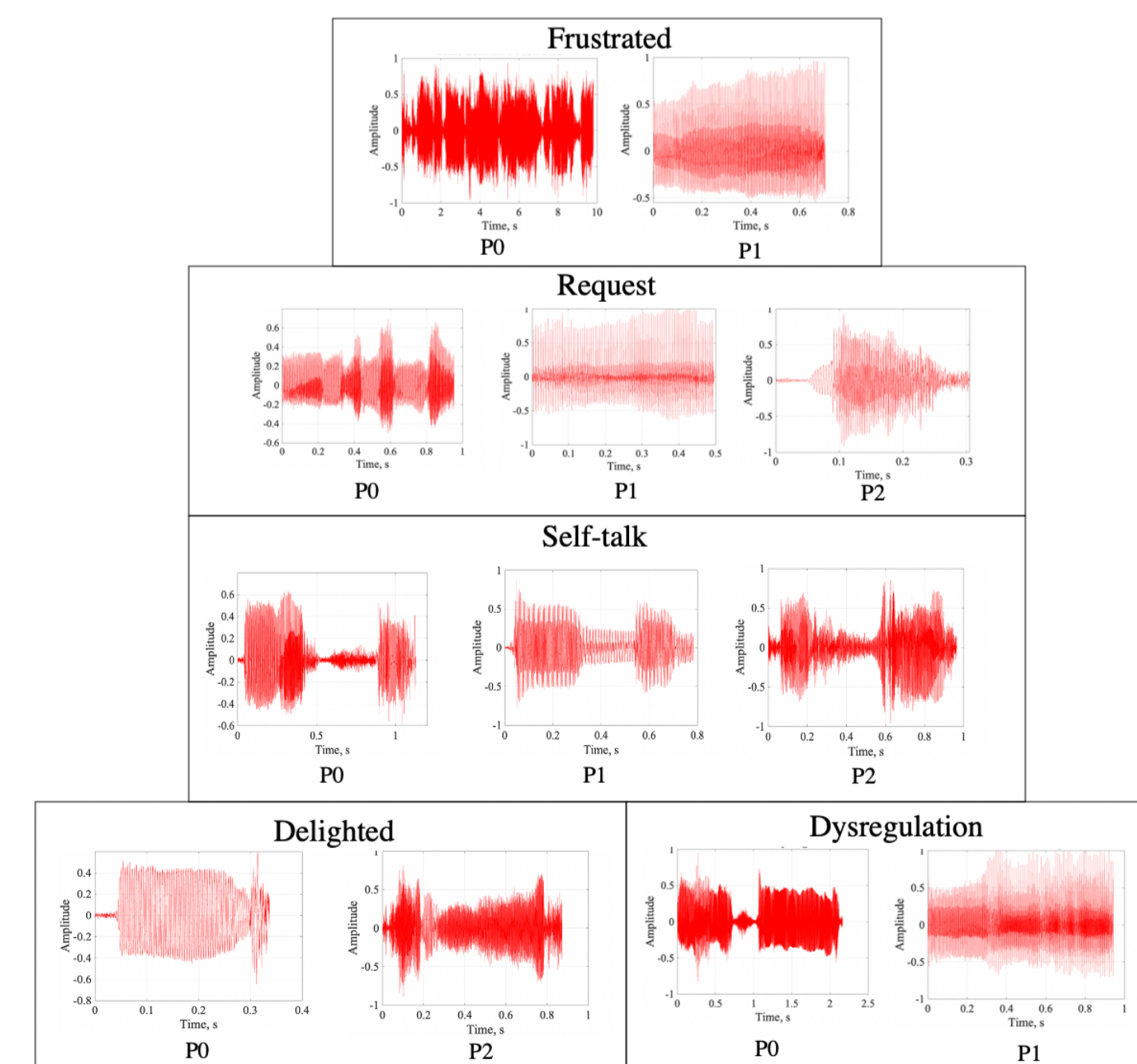
Sample time-domain waveforms of vocalizations in five labeled categories from the three participants (P0, P1, and P2)

Conclusions

- **Distinct vocalization characteristics** relate to individual's **emotional or physical state** and **communicative intent**

- **High within-label variation** due to natural environments

- Opportunity exists to **enhance communicative and affective exchanges** between nv/mv individuals and others

  - We are currently **expanding our data collection** to include more participants

  - We hope to assess the **universal and/or idiosyncratic nature of nonverbal vocalizations** independent of typical speech

  - Dataset will be used to probe scientific questions related to **language and development across ages, genders, and diagnoses**

mit media lab   fluid interfaces   affective computing   Massachusetts Institute of Technology