

Zero-Shot Transfer Learning to Enhance Communication for Minimally Verbal Individuals with Autism using Naturalistic Data

Jaya Narain* & Kristina T. Johnson*, Rosalind Picard, Pattie Maes (*Equal Contribution)

AudioSet indices below match those in [class_labels_indices.csv](https://research.google.com/audioset/download.html) from the Features Dataset found here: <https://research.google.com/audioset/download.html>

As discussed in the text, the three model categories are Laughter, Negative Affect, and Self-Talk. Each positive class contains sounds that were similar (or as similar as possible) to the child's vocalizations for that category. Each negative class contains sounds that might confuse the model because they sounded similar to the child's vocalizations in a different category or appeared frequently in the dataset.

Index	Label Name	Laugh	Neg. Aff	Self-Talk
0	Speech			
1	Male speech			
2	Female speech			
3	Child speech			
4	Conversation			
5	Narration			
6	Babbling			
7	Speech synthesizer			
8	Shout			
9	Bellow			
10	Whoop			
11	Yell			
12	Battle cry			
13	Children shouting			
14	Screaming			
15	Whispering			
16	Laughter			
17	Baby laughter			
18	Giggle			
19	Snicker			
20	Belly laugh			
21	Chuckle, chortle			
22	Crying, sobbing			
23	Infant cry			
24	Whimper			
25	Wail, moan			

 Positive class
 Negative class

[illegible]