



Analysis of Speech Properties of Neurotypicals and Individuals Diagnosed with Autism and Down Syndrome



Mohammed Ehsan Hoque (mehoque@mit.edu)

Media Lab, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139

Abstract

Many individuals diagnosed with autism and Down syndrome have difficulties producing intelligible speech. Systematic analysis of their voice parameters could lead to better understanding of the specific challenges they face in achieving proper speech production. In this study, 100 minutes of speech data from natural conversations between neuro-typicals and individuals diagnosed with autism/Down-syndrome was used. Analyzing their voice parameters indicated preliminary findings across a variety of speech parameters. These findings are currently being used to customize interactive games allowing participants to visualize and control their speech parameters in real time and get live feedback.

Data Collection

The study was set up at the Groden Center, a non-profit school in Providence, RI.

There were six participants in this study: two neuro-typicals (NT), three diagnosed with mild to moderate autism, and one diagnosed with DS.



Fig 1. Experimental data collection setup

The NT partner would arbitrarily pick a topic that s/he believed to interest the other conversation partner (ASD or DS). The conversation then continued as question-answer format with the NT partner taking the lead of the conversation.

Procedures

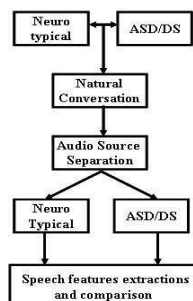


Fig 2. High level diagram.

Speech Features:

Pitch: Minimum, maximum, mean, standard deviation, absolute value, quantile, ratio between voiced and unvoiced frames.

Intensity: Minimum, maximum, mean, standard deviation, quantile.

Formant: First formant, second formant, third formant, fourth formant, fifth formant, second formant / first formant, third formant / first formant

Rhythm: Speaking rate.

Pauses: Number of pauses in an utterance, percent of Unvoiced Frames, maximum duration of Pauses, average duration of pauses, total duration of Pauses etc.

Voice quality: jitter, shimmer, voice breaks (all inter-pulse intervals longer than 16.67 ms are regarded as voice breaks)

Results

Table 1. The optimal feature set of speech samples using combination of (Ranker search technique and Chi Squared Attribute evaluator), (greedy stepwise search technique and CfsSubsetEvaluator), and (Bestfirst search technique with ConsistencySubset evaluator).

Speech features that are similar across three groups (not in any particular order)	Speech features that are dissimilar across three groups (from most to least significance)
Voice quality features (jitter, shimmer), speaking rate, pauses parameters (# of pauses per turn, maximum duration of pauses), maximum duration of pauses, Functions of second formant	Minimum pitch, mean pitch, maximum pitch, mean intensity, mean intensity, max intensity, values of first and third formants, min intensity, energy, bandwidths of first and third formants

Table 2. The comparison of speech properties across Neuro-typicals (NT), Down-Syndrome (DS) and Autism Spectrum Disorder (ASD)

Duration per turn	NT > DS > ASD
Energy	DS > NT > ASD
Appropriate usage of pauses	NT > DS > ASD
Magnitude of rising and falling edges per utterance	NT > DS > ASD
Number of rising/falling edges per utterance	NT = ASD > DS

Screenshot of games



Future work

- 1) Design an experimental setting where participants are encouraged to take turns while accomplishing a task together or playing a game (card games for example).
- 2)The proposed experimental setting should elicit appropriate emotion with unbiased speech samples without requiring someone to take the lead in the conversation.
- 3)Recruit more participants to further validate the reported results.

Acknowledgment

I would like to acknowledge my advisor Prof. Rosalind Picard for her valuable guidance with this work.

This research was partially supported by NSF (IIS HCC – 0705647) and Nancy Lurie Marks Family Foundation (NLMFF).