Entrainment Analysis for Assessment of Autistic Speech Prosody Using Bottleneck Features of Deep Neural Network

Keiko Ochi(Kyoto University), Nobutaka Ono (Tokyo Metropolitan University), Keiho Owada, Miho Kuroda ,
Shigeki Sagayama (University of Tokyo), Hidenori Yamasue (University of Tokyo /Hamamatsu University School of Medicine)

What is autism spectrum disorder (ASD)?

- Deficits in **social communication and interactions** (facial expressions, eye gazes, pragmatics, **speech prosody**)
- 1/54 children [Maenner2020+]
- **Problems of current assessment:**
  - **-Subjective rating**
  - -Difficult to use repeatedly

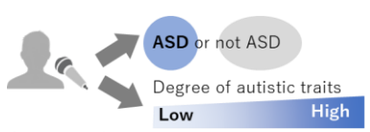
E.g., the **Autism Diagnostic Observation Schedule (ADOS)**

Time-consuming

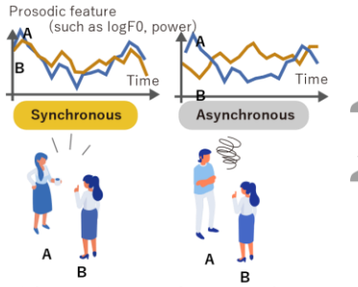
- Needs for repeated assessment to develop of novel treatment/medication

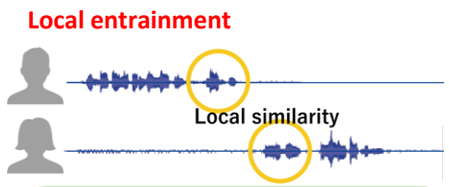
Quantification of speech in people with ASD

- Why **speech signals?**
  - -easy to obtain
  - -can be utilized in automated diagnosis/assessment



- In **conversational speeches** in speakers with ASD:
  - **longer turn-taking gaps and pausing** [Heeman+2010] [Bone+2016]
  - **less global entrainment** with their interlocutors in terms of
    - speech rate [Wynn+2018]
    - F0/intensity entrainment [Ochi+2019]
- However, the local entrainment of people with ASD is still unclear.





High entrainment relates to task success

When entrainment occurs, statistics (i.e., mean and standard deviation SD) of prosody are similar.

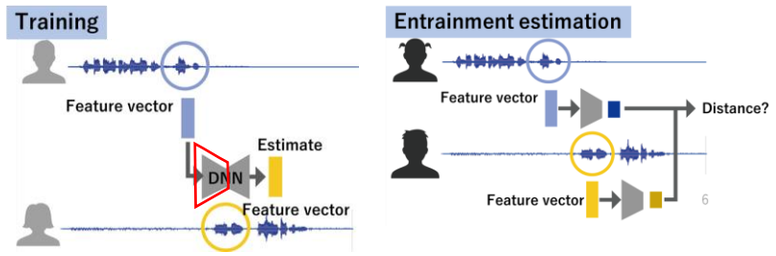
Purpose of this study

Quantify **the local entrainment for automated assessment**
- entrainment between patients and their interlocutors

- aiming at a novel, **easy-to-use assessment** method for ASD,
- using conversations in **semi-structured interviews**
  - to control the contents of the dialogues.
- Analyzing prosodic/acoustic features
  - **just before and after turn-takings**

Related work: neural entrainment distance [Nasir+'18]

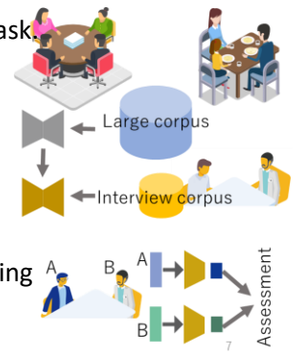
- Train an hourglass-shaped DNN
  - predict the next turn's speech features from that of the preceding turn



- Regard the **distance between the two bottleneck features** as the **degree of entrainment**.

Proposed method for automatic assessment

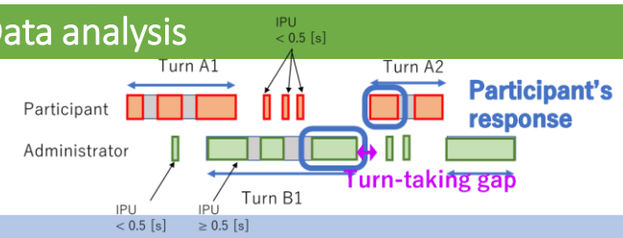
- Quantify entrainment based on [Nasir+'18]
  - Changes: input features according to the task
- Pre-train the DNN using a task-nonlimited corpus without considering ASD.
- Fine-tune with semi-structural conversation
  - Use data from people with typical development (TD) (control group)
  - Fit the DNN to the task
- Access ASD by bottleneck feature vectors
  - Embedding the local entrainment embedding

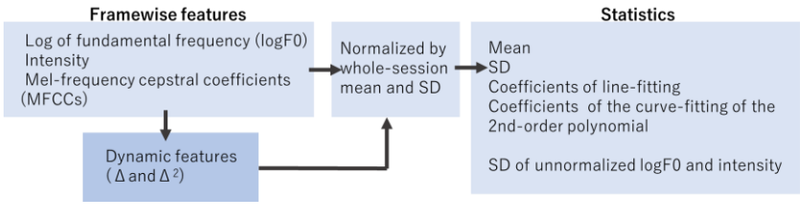


Speech datasets

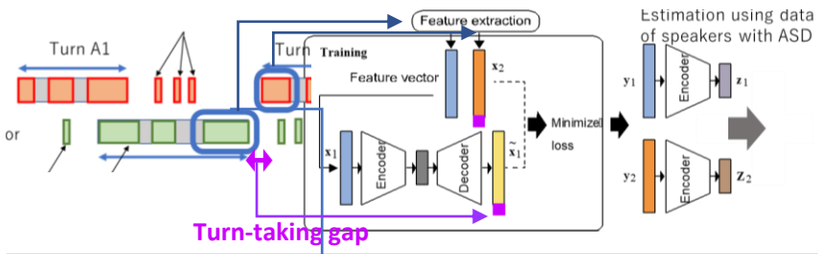
- **Corpus of Everyday Japanese Conversation (CEJC) (monitor version)**
  - 50-hour audio data
  - 118 conversational situations among 3-9 people
- **Semistructured-interview conversation dataset**
  - ADOS Module 4 administration
    - as a part of a **clinical trial of medicine** (Oxytocin nasal spray) in the hospital of the University of Tokyo
    - **before the medication** of oxytocin/placebo
    - Activity 7 (**Questions and answers about emotion**)
    - 82 recordings (101-389 sec)
    - Participants: 65 male adults with ASD and 17 controls

Data analysis





- **DNN configuration for entrainment qualification:**
  - -5 fully connected layers (316-64-16-64-317)
  - -Pre-training: 39152 samples (90% for training, 10% for validation)
  - - Fine-tuning: 4438 samples (95% for training, 5% for validation)
- **Automated assessment of ASD**:
  - Estimate ADOS scores
  - Use support vector regression (SVR)
  - Input features: the centroid and SD of the difference of the bottleneck feature vector
  - Select feature dimensions by forward feature selection (FFS) method
    - by adding the speech features of our previous study



Results

Table: Correlations and mean absolute error (MAE) (shown in parentheses) between the estimated and the observed ADOS score in the leave-one-out cross-validation

Method	Reciprocity facial/verbal interaction	Communication Nonverbal interaction	Repetition Repeating a certain behaviors
Baseline	0.59 (1.23)	0.49 (0.96)	0.18 (1.23)
Without fine-tuning	0.60 (1.23)	0.59 (0.90)	0.49 (0.61)
Proposed (with fine-tuning)	**0.70 (1.18)**	0.62 (0.84)	0.49 (0.61)

Proposed features provided the best performance for every three categories of ADOS score.

Discussions and conclusions

- High assessment performance with local entrainment
- Successfully estimated score of the highest (most autistic) participant
- Overestimated score of the lowest-score (less autistic) participant
  - Speech prosody was rated high by human
  - Other sub-items (visual information) were rated low
  - ⇒The prosodic characteristics affected the estimation.
- Future works:
  - applying the automatic evaluation to the therapies such as computer-assisted social skill training