TAPAS
Training Adapted Personalised Affective Social Skills with Cultural Virtual Agents

ANR AGENCE NATIONALE DE LA RECHERCHE

JST Japan Science and Technology Agency
CREST Symbiotic Interaction

Multimodal Dataset of Social Skills Training in Natural Conversational Setting

Takeshi Saga[1], Hiroki Tanaka[1], Hidemi Iwasaka[2],

Yasuhiro Matsuda[2], Tsubasa Morimoto[2], Mitsuhiro Uratani[2],

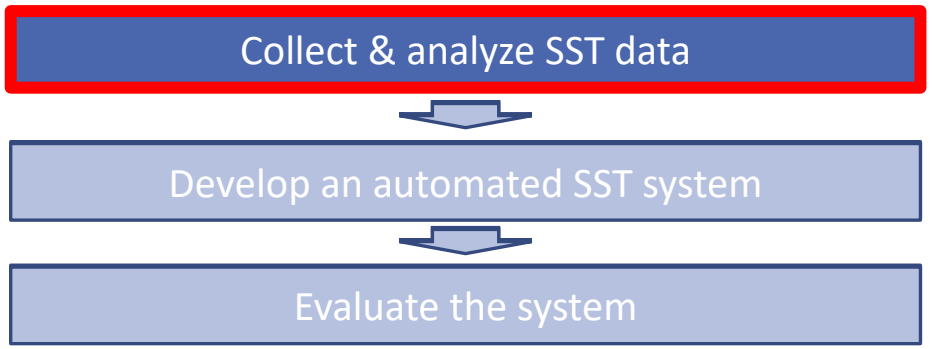
Kosuke Okazaki[2], Yuichiro Fujimoto[1], Satoshi Nakamura[1]

[1] Nara Institute of Science and Technology, [2] Nara Medical University

Our research background

- **Long-term objective**
  - **to develop methods and tools to reduce social stress** in everyday situations (e.g. public speaking, social communication)
- **Method**
  - Social Skills Training (SST), etc.
- **Target population**
  - Healthy people
  - Schizophrenia
  - Autism Spectrum Disorder (ASD)

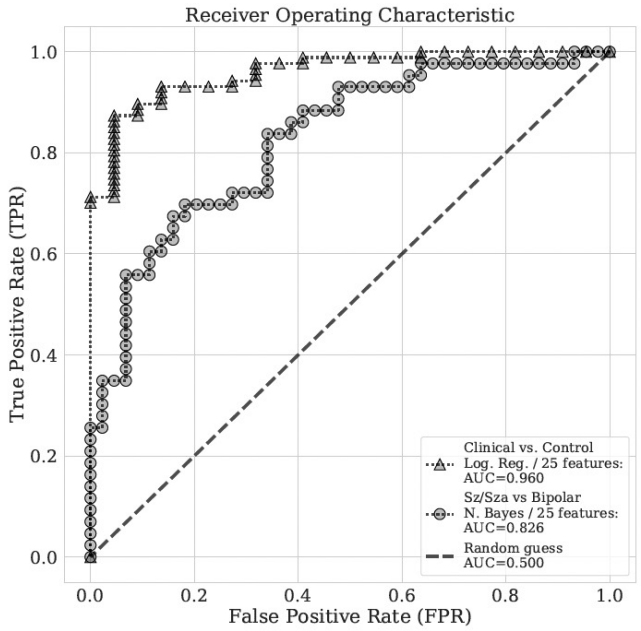
- **Research flow**



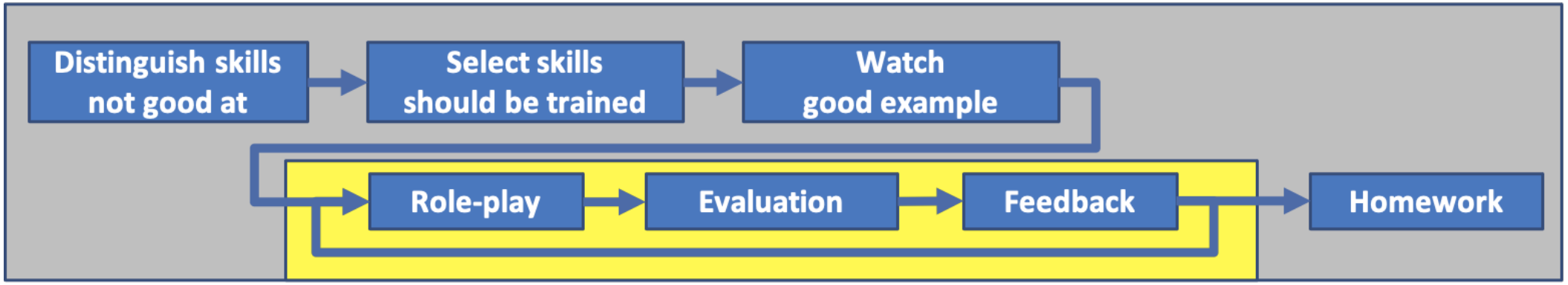
- **What we'd like to know**
  - Behavior analysis in SST
    - Disease-specific difference in SST
    - User-specific difference in SST

Related work

- **Voleti+2019**
  - Classification using language features
    - Semantic coherence
    - Lexical diversity
    - Lexical density
    - Syntactic complexity
  - Task: Social Skills Performance Assessment
  - Target disease
    - Schizophrenia, Schizoaffective disorder
    - Bipolar
    - Control
  - Classification AUC
    - 0.960 for Clinical vs Control classification
    - 0.826 for Sz/Sza vs Bip classification



Social Skills Training (SST)



A training framework used in psychiatric rehabilitation programs

Especially effective for people with difficulties related to mental illnesses or developmental difficulties, such as Schizophrenia(SZ) or Autism Spectrum Disorder(ASD)

Bellack method [Bellack+2004] is naturalistic style SST setting

Overview of this research

Problem

No research on behavior analysis in SST, especially for SZ and ASD

Key idea

1) Collect data to analyze behavior difference in SST

2) To maximize the training effect, disease-specific method is needed

Our contribution

1) Created human-human natural SST dataset with Control/ASD/SZ groups

2) Analyzed with Control/SZ classification model using the dataset

Dataset

Dataset — setting

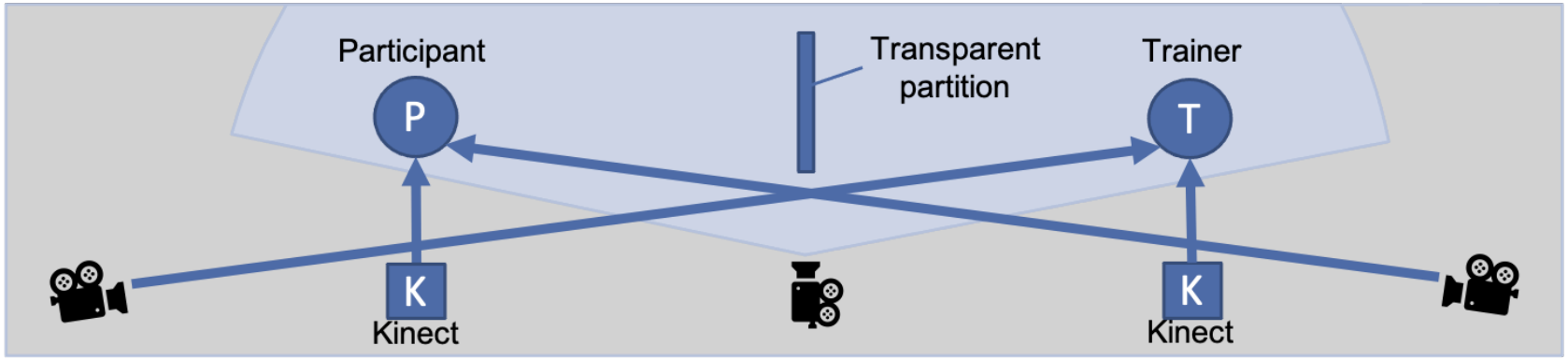
Video camera for facial expression

Azure kinect for body movement

Transparent partition to prevent Covid-19 infection

Four role-play situations for each participant

- Express positive feeling, listen to others, ask a favor, refuse an offer



Dataset — diagnosis, assessment

Clinical Assessment

- Social skills — FEIT, Kiss-18, Singelis
- ASD related — SRS-2, ADOS-2
- Schizophrenia related— BACS-J, PANSS

Note: ASD, Schizophrenia
- Recruited based on Psychiatrist diagnosis

- **3rd-party psychiatrist evaluation**:5-point Likert scale, n=2, kappa=0.84
  - For participants — seven components
    - e.g. eye-contact, facial expression
  - For trainers — four components
    - e.g. appropriateness of positive feedbacks, suggestions for improvement

Classification

BEHAVIOR DIFFERENCE BETWEEN DISORDERS IN SST

Classification — method

Subjects — Schizophrenia and Control

Model — random forest

Preprocessing

- Separate video and audio files into Role-play or Feedback segment files
- Eliminate vocal overlap from audio
- 48 audio segments for Schizophrenia , 46 audio segments for Control

Input features — 20 audio related and 21 face related (by Praat, OpenFace)

- Three different condition: audio-only, visual-only, audio-visual

Training — leave-one-subject-out cross validation

Classification — result & discussion (1)

Feature	Accuracy	Precision	Recall	F1-score
A	0.868	0.902	0.872	0.887
V	0.382	0.476	0.488	0.422
A+V	0.804	0.823	0.851	0.836

A : Audio feature-set
V : Visual features-set

From the result ...

- Audio-only feature-set achieved the best scores.
- By adding visual features the accuracy was dropped.

Characteristic differences in the audio features

Classification — result & discussion (2)

Feature name	F3 BW	F2/F1 Mean	F1 BW
Importance	0.231	0.089	0.087

From the top-three feature importance ...

- All of them were audio features
- Bandwidth of F3 (F3 BW) is clearly higher than others

Wide formant bandwidth induces a significant reduction of the vowel identification rate [Dubno+1987]

F3 BW indicates the less vowel intelligibility in Schizophrenia group

Schizophrenia with strong positive symptom have more complex articulatory coordination [Siriwardena+2021]

Conclusion

Summary

- We collected natural SST dataset with SZ and ASD people
- We experimentally trained a classification model with 88.7% F1 score

Limitation

- Gender imbalance of trainers
- Didn’t consider text contents

Future work

- Classification with ASD/Schizophrenia/Control
- Add more modalities (text, body)
- Try sequential model to consider features related to synclonization

Supplemental

Comparison

		Voleti+2019	Ours
Dataset	Task	Social Skills Performance Assessment	Social Skills Training
	Classification target	Sz,Sza/Bip/Control	Sz,ASD,Control
Analysis	features	text	audio/visual

Our research project: TAPAS

- **Work packages**
  - WP1: Situation & immersion
  - WP2: Online measuring & feedback
  - WP3: Virtual agent platform
  - WP4: Gathering therapy data
  - WP5: Evaluation

- **Objective**
  - to develop methods and tools to reduce social stress in everyday situations (e.g. public speaking, social skills training)

- **Target population**
  - public speaking
  - Social Anxiety Disorder (SAD)
  - Schizophrenia
  - Autism Spectrum Disorder (ASD)

Dataset — clinical assessment

	Group	Control-adult	ASD-adult	Schizophrenia
Perception of facial expression	FEIT	✓	✓	✓
General social skills scales	Kiss-18	✓	✓	✓
General social skills scales	Singelis	✓	✓	✓
Metric for ASD traits	SRS-2	✓	✓	
Metric for ASD traits	ADOS-2		✓	
Metric for cognitive functionality	BACS-J		✓	✓
Metric for Schizophrenia symptoms	PANSS			✓

Features

Feature name	Description
F0, F1, F2, F3 Mean	Mean frequency of F0, F1, F2, F3
F0, F1, F2, F3 SD	Standard deviation of F0, F1, F2, F3
F0 Min, Max	Minimum and Maximum F0 frequency
F0 range	Range (Max - Min) of F0
F1, F2, F3 BW	Average bandwidth of F1, F2, F3
F2/F1, F3/F1 Mean	Mean ratio of F2-F1 and F3-F1
F2/F1, F3/F1 SD	Standard deviation of F2/F1 and F3/F1
Int range	Range (Max - Min) of intensities
Int SD	Standard deviation of vocal intensity

Feature name	Description
Pose_[Rx,Ry,Rz]_std	Standard deviation of head rotation around (X:pitch,Y:yaw,Z:roll) axis
Smile_ratio	Ratio of Smiling frames
AU01	inner-brow raiser
AU02	outer-brow raiser
AU04	brow lowerer
AU05	upper-lid raiser
AU06	cheek raiser
AU07	lid tightener
AU09	nose wrinkler
AU10	upper-lip raiser
AU12	lip-corner puller
AU15	lip-corner depressor
AU17	chin raiser
AU20	lip stretcher
AU23	lip tightener
AU25	lips parter
AU26	jaw dropper
AU28	lip sucker
AU45	blinker