Non-verbal Cognitive Skills and Autistic Conditions: An Analysis and Training Tool

Hiroki Tanaka, Sakriani Sakti, Graham Neubig, Tomoki Toda, Nick Campbell, Satoshi Nakamura
Graduate School of Information Science
Nara Institute of Science and Technology
Ikoma-shi, Nara, Japan.
hiroki-tan, ssakti, neubig, tomoki, nick, s-nakamura@is.naist.jp

***Abstract*—The number of People who have trouble with social skills and communication is now greater than ever for a variety of reasons. Autism spectrum conditions are an extreme example of these traits. Our objective is to measure these autistic traits automatically, and enable people with social and communication difficulties to improve social and communication skills for use in the real world. This paper examines the relationship between non-verbal cognitive skills and position on the autism spectrum among members of the general population, and pre- and post-learning results were examined to find the effects of the training. The results showed an improvement after a 20-minute learning session, indicating that training could help enhance non-verbal cognitive skills for members of general population.**

***Keywords*-non-verbal information, autism spectrum condition, partner information, mobile application**

I. Introduction

An article in the September 2005 issue of the Hiragana Times states that "the generation raised by the one-way information provided on TV is poor in communication." People who have trouble with social skills and communication have recently been increasing due to this and other reasons. These difficulties cause trouble to relating to other people [1]. Both identifying the degree of these difficulties and developing a learning tool for social and communication skills are necessary.

Autism is a set of neurodevelopmental conditions characterized by social interaction and communication difficulties, as well as unusually narrow, repetitive interests [2]. The diagnosis criteria of autism includes a "marked impairment in the use of nonverbal behaviors, such as eye-to-eye gaze, facial expression, body posture, and gestures to regulate social interaction [14]." Fujisaki [15] uses the term "non-verbal" to refer to not only emotion, but also partner information, intention, situation, age, sex and other factors.

One of the central psychological themes in autism is empathizing. Empathizing is a set of cognitive and affective skills we use to make sense of and navigate the social world. The cognitive component of empathy is also referred to as theory of mind [3]. It is well established that emotion recognition and mental state (non-verbal) recognition are core difficulties in people with Autism Spectrum Conditions (ASC). Such difficulties have been found across different sensory modalities, both visual and auditory. Neuroimaging studies of emotion recognition from faces reveal that people with ASC show less activation in brain regions central to face processing, such as the fusiform gyrus [4]. There is also evidence of reduced activation in brain areas that play a major role in emotion recognition, such as the amygdala, when individuals with ASC process socioemotional information [4], [5].

One of the factors influencing the ability to empathize is the severity of ASC. Autism is a spectrum condition [6], that has a broad range of clinical characteristics ranging from mild to severe. There are several methods for measuring a person's position on the autistic spectrum. For example, the Autism Spectrum Quotient (AQ) [7] is a self-administered screening measurement that can be used for children from 4 years of age through to adulthood, and individuals score in the range of 0-50 by answering a total of 50 statements. However, a high AQ score alone is not a reason to be referred for a diagnosis; there also has to be evidence that the person is suffering. The AQ measures how many autistic traits an individual shows, and can be used across the general population, not only with people who are suspected of having ASC.

Several papers note the fact that among the member of general population, autistic conditions are widely distributed along a spectrum. It is reported that autism occurs more often in families of physicists, engineers, and mathematicians, and there is a link between engineering and autism [8], [9]. Likewise, the AQ was tested among 840 students in Cambridge University, and the result showed that scientists scored higher than both humanities and social scientists. This confirmed the association between science/maths skills, and autistic conditions.

There are large amount of learning tool for autism [10]. Golan et al. suggested the Mindreading DVD, which enable adults with autism to learn mental state recognition. The improvement of emotion recognition skills was indicated through several months training. However other generalization levels (questions not include in training) were still difficult [11], and these typically do not include non-verbal speech signals. Beukelman [12] also claims that we need to document the successful strategies that allow Augmentative and Alternative Communication (AAC) [13] users to communicate background messages that convey emotion and mood. Such emotion and mood is expressed in non-verbal information.

Taking into consideration the issues mentioned above, we attempt to; first make clear relationship non-verbal cogni-

TABLE I
FACTOR ANALYSIS USING THE PROMAX ROTATION METHOD [16].

AQ no.	Statement	Factor loadings				
		1	2	3	4	5
	[intention, interest]					
45	I find it difficult to work out people's intentions.	1.308		-0.294		-0.191
35	I am often the last to understand the point of a joke.	0.687	-0.12		0.143	-0.109
15	I find myself drawn more strongly to people than to things.	0.613	0.263		-0.117	0.112
1	I prefer to do things with others rather than on my own.	0.571	0.436	0.138		
	[polite, new friend]					
22	I find it hard to make new friends.		0.869	0.114		0.187
7	Other people frequently tell me that what I've said is impolite, even though I think ...		-0.722			0.282
27	I find it easy to "read between the lines" when someone is talking to me.	0.159	-0.701		0.124	-0.129
47	I enjoy meeting new people.	0.161	0.524	-0.147	0.124	0.153
26	I frequently find that I don't know how to keep a conversation going.		0.515		0.189	-0.243
	[social place and situation]					
13	I would rather go to a library than a party.	-0.19	0.159	1.079		
48	I am a good diplomat.	-0.117	-0.225	0.734	0.201	0.768
18	When I talk, it isn't always easy for others to get a word in edgeways.	0.364	0.314	0.396	-0.179	
11	I find social situations easy.	0.281	-0.29	0.372		
	[chit-chat, feeling]					
31	I know how to tell if someone listening to me is getting bored.			-0.325	0.833	
17	I enjoy social chit-chat.			0.366	0.735	0.108
38	I am good at social chit-chat.	-0.212	0.128	0.309	0.531	-0.248
44	I enjoy social occasions.	0.384		0.175	0.492	
36	I find it easy to work out what someone is thinking or feeling just by looking at ...	0.282		-0.213	0.475	0.219
	[others]					
33	When I talk on the phone, I'm not sure when it's my turn to speak.	-0.378	0.365		0.135	0.851
39	People often tell me that I keep going on and on about the same thing.	0.358	-0.283	-0.144	-0.317	0.552
	SS loadings	3.125	3.085	2.591	2.283	2.097
	Cumulative Var	0.156	0.31	0.44	0.554	0.659

TABLE II
CORRELATION COEFFICIENT BETWEEN FACTORS AND SOCIAL AND COMMUNICATION SKILLS (***INDICATES P<0.001, **INDICATES P<0.01, AND *INDICATES P<0.05 BY T-TEST).

			1	2	3	4	5
	Social and communication skill	1.00	0.73**	0.60***	0.79***	0.67***	0.08
1	Intention, interest		1.00	0.31	0.55**	0.25	-0.004
2	Polite or impolite, new friend			1.00	0.27	0.32	0.07
3	Social place and situation				1.00	0.44*	-0.11
4	Chit-chat, feeling					1.00	-0.27
5	Others						1.00

tive skills and autistic conditions, second; develop a training method of social and communication skills.

To evaluate the first goal, we evaluate the adult AQ scores to confirm the non-verbal factors contributing to social and communication skills, which include, but are not restricted to emotion. To achieve second goal, and develop a mobile application reflecting the result of this analysis of a AQ tendencies. The mobile application allows users to measure autistic traits automatically, and enables people with social and communication difficulties to improve non-verbal cognitive skills for use in the real world.

II. ASSESSMENT OF COMMUNICATION SKILLS

Non-verbal information includes various factors (e.g., eye contact, intention, gesture, and sex). The objective of this section is to confirm the important non-verbal factors contributing to communication skills as measured by using AQ. To do so, we use Factor analysis, which is commonly used to elucidate the factors contributing to scores on a psychometric test. To collect data, we first asked 21 Japanese students to take the English version of the AQ to measure two of the original five areas: social and communication skills (with a total of 20 statements).

The Cronbach's coefficient alpha is commonly used as a measure of the internal consistency or reliability of a psychometric test score for a sample of examinees. It is calculated with following formula:

$$\alpha = \frac{m}{m-1}\left(\frac{\sum_{j=1}^{m}\sigma_j^2}{\sigma_x^2}\right) \tag{1}$$

where m is the number of components (m-items), σ_x^2 is the variance of the observed total test scores, and σ_j^2 is the variance of component j for the current sample of persons. This resulted in Cronbach's coefficient alpha value of 0.73 (> 0.7), indicating that the test is reliable.

Next, we perform a factor analysis to determine several important factors for social and communication skills based on the AQ. Based on using principal component analysis (PCA) and the chi-square value we finally set 5 factors. Table I shows the loadings and the proportion of variance from the first factor to the fifth factor (the cumulative proportion of variance is up to 65%). Each individual factor's contribution ratio is not high, even for the first factor. Next, we perform an analysis with the promax method [16], which is an alternative non-orthogonal (oblique) rotation method that is effective when there are highly correlated factors. This reveals the following:

1) the first factor is largely related to intention and interest.

2) the second is related to politeness or impoliteness as well as new friends.
3) the third is related to social places and situations.
4) the fourth is related to chit-chat and feelings.
5) the fifth is other factors.

To confirm the degree to which each factor is effective for evaluating communication and social skills, we calculate Pearson's r value between each factor's total score. Table II reveals that the first five factors are sufficient to measure social and communication skills. As a result we selected the first two factors (intention & interest, and politeness/impoliteness & new friends) as non-verbal information. Finally these represent intention and partner information.

III. Classification of Natural Speech

In this section, we performed classification of natural speech data according to previous section. The categorized utterances can be used to measure and learn non-verbal cognitive skills.

A. Natural conversational speech corpus

The FAN subset of JST/CREST Expressive Speech Processing (ESP) corpus [17] was recorded over a period of five years, and consists of over 600 hours of every-day conversational speech collected from a female volunteer, who used a high-quality head-mounted microphone to record her speech to a small mini-disc recorder. This corpus features a large amount of speech from various situations, including simple, repetitive and unstructured talk that shows how people actually speak in everyday situations. We prepared a total of 5,367 short utterances from the FAN database following previous work [18].

B. Communication skill categorization

Based on the result of the factor analysis described in Section 2, we decide to use the first two factors plus another factor for content of conversation, which is essential for speech communication. The resulting axes are content of the utterance, partner information, and intention. The utterances were classified into one of the 3 types of categories by 3 Japanese students (male, ages: 23 and 24). The final total number of content of the utterance was 27. The final categories of partner information was "friend" and "teacher", and intention was "derisive", "social", and "friendly." The categorization procedure is following.

The numbers of categories of partner information and intention in each axis were determined subjectively, bottom up, and only utterances that the 3 students all agreed upon were left in the database. As a result, utterances (content: 2, partner: 3, intention: 6) are chosen. For partner and intention, the annotators separated 60 randomly chosen utterances into categories were family, teacher, and friend. For these three categories, the agreement value was only 50%, which indicates low agreement. To resolve this problem we merged the family and friend categories to resolve this problem, as the error rate between these two categories was the highest. In terms of intention there were 6 categories from bottom-up, and Cohen's multi-Kappa statistics were 0.32, which indicates low agreement. Thus we calculated Euclidean distance between the clusters (which similar to error rate), and employed re-clustering. As a result, Cohen's multi Kappa statistics rose to more than 0.6. The final three categories were: derisive, social, and friendly.

IV. Mobile Application

Finally we developed a mobile application named NOCOA, which reflects the above result of overall AQ tendencies and classification of short utterances.

A. Voice Conversion

The speech used in NOCOA is converted to sound more like a child's voice considering that our final target is autistic children, and to protect FAN's privacy. We used the software MacSynthTransformer [19], which allows for changing the pitch and envelope of speech. 4 Japanese students (3 Male, 1 Female) listened to the two varieties of speech (original speech and converted speech), and confirmed that the speech quality was not reduced.

B. Facial images

Next we prepared facial images for each category. As described in Section 3, we chose three axis: 27 types of contents of the utterance, 3 types of intention or interest of talk, and 2 types of partner information. Both actual pictures (chosen via yourstock.com) and illustrations were prepared for each category, and the use of pictures or illustrations can be chosen by the user.

C. Structure

This subsection explains two modes of NOCOA.

1) Listening mode: In listening mode, users touch the screen to choose the content, choose from two types of partner information, and then choose from three types of intention. If there is no available sound candidate, the photos will be blank. Finally the user can see the result they chose on the play screen, and can listen to the appropriate sound. The maximum number of sounds in each category is 4, and the sound is played randomly.

2) Test mode: NOCOA also has a test mode, which is able to measure users' intention and partner information cognitive skills. The user listens to the voice, and then chooses the appropriate face and partner. The test mode score is calculated by using agreement in each category with the general population. The intention category's score penalty for mistakes between derisive and social is higher than for those between social and friendly because these are critical misses in a social situation. In both partner information and intention the maximum score of each question is 5. The test mode score is calculated after answering 10 questions, so 100 is the best score. The 10 question set is chosen at random each time.

Here, computer-based intervention used drawings of photographs for training, rather than more lifelike stimuli. This

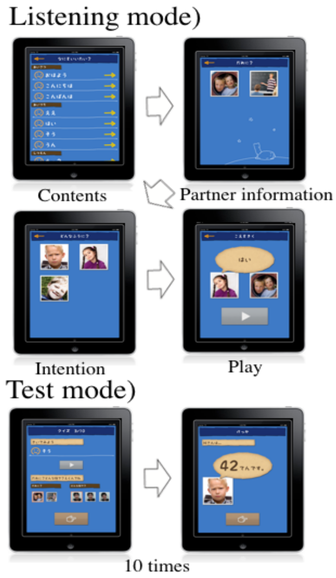

Fig. 1. Two modes of NOCOA, listening mode and test mode. Both modes were developed systematically.

might have made generalization harder than if more ecologically valid stimuli were used. Thus test mode also has three generalization levels:

1) Closed data: testing was performed using voices that were included in the listening mode but faces were presented using a different person.
2) Open data: faces and voices were not included in the training, but the content was the same as in the training.
3) Long sentences: faces and voices were not used in the training, and the content was not included in the listening mode, because the main utterances used in training are short.

V. Experiment: an Evaluation

A. Experimental Setting

We performed an experimental evaluation of the correlation between AQ score and test mode score in members of the general population. We perform this evaluation because our tool was developed for people who have difficulties with social and communication skills to measure their non-verbal cognitive skills and to systematically learn how to identify non-verbal information.

The procedure of the experiment is as follows: 19 Japanese participants were recruited (mean age is 25.0, 18 males and 1 female). They came to the laboratory one by one, and took the AQ test. After finishing, we checked the understanding of the concept of facial images mentioned in section 4.2. We confirmed that all participants did not have difficulty in understanding the concepts. Then, participants took generalization level 1 (closed data) of test mode on NOCOA two times, and the average score of two trials was calculated.

We also tested efficacy of listening mode with several Japanese students (training group) who scored below average (mean age is 23.0). They used listening mode for 20 minutes, and the control group waited for the same 20 minutes. After 20 minutes both groups used test mode with the three generalization levels.

B. Experimental Results

First, we measured relation between test mode score and AQ, the correlation coefficient between AQ and averaged test mode score was 0.70 (see Figure 2). This reveals that large variations in the ability to recognize non-verbal and partner information exist in the general population, and is significantly related to autistic traits. Note that despite the fact that the participants had not been diagnosed with Asperger syndrome or high-functioning autism who have average or above average IQ, their range of AQ scores was wide and well correlated with test mode score.

We also tested efficacy of listening mode with two Japanese students who scored below average (mean age is 23.0) and participated in training. Figure 3 shows that after using listening mode for 20 minutes, their score also improved above 10 points in the case of generalization level one. As a result of training we found they maintained high score in both open and long utterances (Figure 4).

Here, we have to consider other factors to verify that the result is reliable. Because the FAN corpus is recorded by a person with a Kansai accent, which types of dialect in Japan, we calculated the averaged test mode score between people with Kansai accent and people without Kansai accent in participants. The results showed with Kansai accent participants achieving 82.1 (11 people), and not Kansai accent participants achieving 82.9 (8 people).

VI. Conclusion

In this paper we confirm the relationship between non-verbal cognitive skills and AQ by using speech output with visual hints, and examine prospective intervention through teaching non-verbal information, intention and partner information. Classification of utterances is based on the AQ, and our analysis revealed that it is an effective way to measure social and communication skills. According to factor analysis, we confirmed three important axes. Previous reports mention basic or complex emotions [20], [21], [11] or involvement [22] but not partner information. The number of categories of partner information and intention in each axis was determined subjectively and bottom up. We conducted a subjective experiment with members of the general population, and confirmed that this tool is useful. As a result of experiment, correlation between AQ score and test mode score was 0.70 (p-value $< .01$) for 19 Japanese adults. This shows that ASC severity is significantly related to test mode score even in the Japanese adult group. It also reveals that in the general population, where the range of AQ scores was wider, the more autistic traits one possesses results in recognition of non-verbal information being more difficult. In addition several Japanese students had difficulty distinguishing utterances compared to other members. However their test mode score was improved

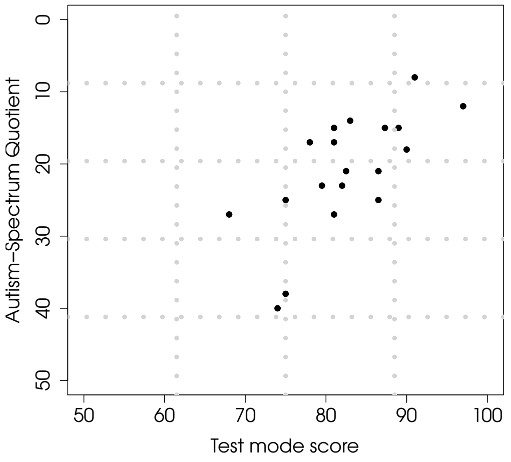

Fig. 2. Correlate between test mode score and AQ score (Pearson's r value is 0.70, p-value $<$ .01) by 19 Japanese adults.

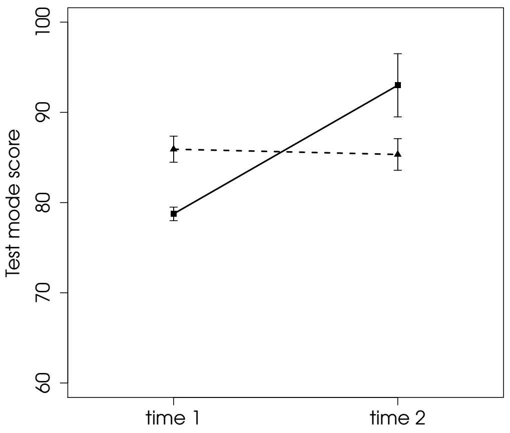

Fig. 3. The test mode scores between before 20 minutes training (time 1) and after the training (time 2) with standard error bar. The purple line shows control, and the red line shows training group.

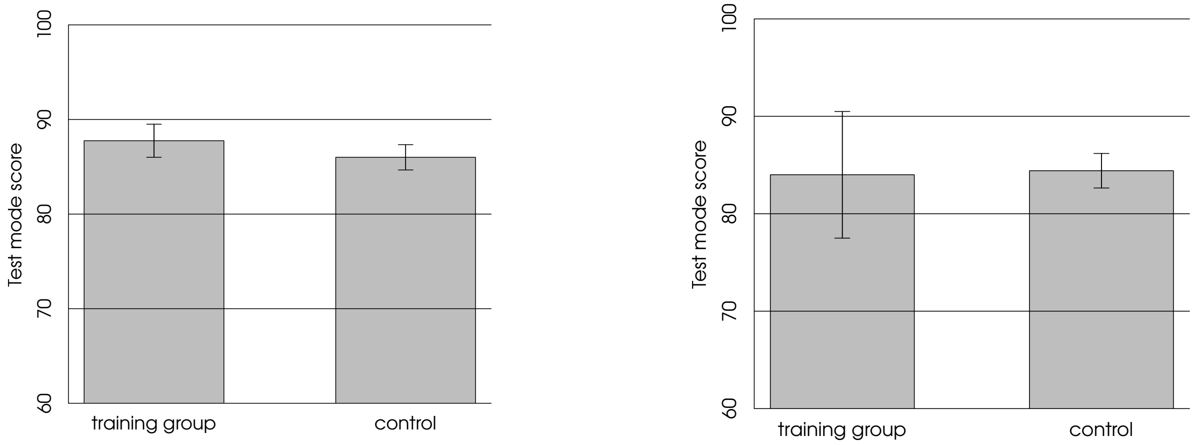

Fig. 4. The score between training group and control. A left figure shows the result in generalization level two, and a right one shows the result in generalization level three.

by using listening mode for 20 minutes. They also maintained high scores even in unseen open questions and long sentences.

Our goal is further improve towards supporting real communication. Though this paper presented visual hints as static graphical images, to more accurately simulate real-world situations, we will try to use movie data. One challenge is how to collect natural data.

We have been distributing NOCOA through the Apple AppStore for educational use since February 2012.

(see http://itunes.apple.com/ph/app/nocoa/id501936653?mt=8)

Acknowledgment

The author thank Dr. Hidemi Iwasaka, for his advice and support.

References

[1] BARON-COHEN S. 2008. Autism and Asperger syndrome. Oxford University Press, USA.

[2] KANNER L. 1943. Autistic disturbances of affective contact. Nervous Child 2: 217-250.

[3] ASTINGTON J.W., P.L. HARRIS & D.R. OLSON 1990. Developing theories of mind. Cambridge Univ Press.

[4] CRITCHLEY H.D., DALY E.M., BULLMORE E.T., WILLIAMS S.C.R., VAN AMELSVOORT T., ROBERTSON D.M., ROWE A., PHILLIPS M., MCALONAN G. & HOWLIN P. 2000. The functional neuroanatomy of social behaviour. Brain 123: 2203-2212.

[5] ASHWIN C., BARON-COHEN S., WHEELWRIGHT S., O'RIORDAN M. & BULLMORE E.T. 2007. Differential activation of the amygdala and the social brain during fearful face-processing in Asperger syndrome. Neuropsychologia 45: 2-14.

[6] WING L. 1996. Autistic spectrum disorders. Bmj 312: 327.

[7] BARON-COHEN S., WHEELWRIGHT S., SKINNER R., MARTIN J. & CLUBLEY E. 2001. The Autism-Spectrum Quotient (AQ): evidence from

Asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. Journal of Autism and Developmental Disorders 31: 5-17.

[8] BARON-COHEN S., WHEELWRIGHT S., STOTT C., BOLTON P. & GOODYER I. 1997. Is there a link between engineering and autism? Autism-London- 1: 101-109.

[9] BARON-COHEN S., BOLTON P., WHEELWRIGHT S., SHORT L., MEAD G., SMITH A. & SCAHILL V. 1998. Autism occurs more often in families of physicists, engineers, and mathematicians. Autism 2: 296-301.

[10] MOORE D., MCGRATH P. & THORPE J. 2000. Computer-aided learning for people with autism—a framework for research and development. Innovations in Education and Teaching International 37: 218-228.

[11] GOLAN O.F.E.R. & BARON-COHEN S.I.M.O.N. 2006. Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. Develop. Psychopatholy. 18:

[12] BEUKELMAN D. 1989. There are some things you just can ’ t say with your right hand. Augmentative and Alternative Communication 5: 257-258.

[13] BEUKELMAN D. & MIRENDA P. 2005. Augmentative and alternative communication.

[14] AMERICAN PSYCHIATRIC ASSOCIATION. 1994. The diagnostic and statistical manual of mental disorders, IV . Washington, D.C.: American Psychiatric Association.

[15] FUJISAKI H. 1997. Prosody, models, and spontaneous speech. Computing Prosody.

[16] TANAKA Y. & WAKIMOTO K. 1983. Methods of multivariate statistical analysis. Gendai Sugoku, Tokyo.

[17] Expressive speech processing corpus: www.speech-data.jp

[18] CAMPBELL N. 2006. Conversational speech synthesis and the need for some laughter. IEEE Transactions on Audio, Speech, and Language Processing. 14: 1171-1178.

[19] ROEBEL A. 2010. A shape-invariant phase vocoder for speech transformation. Proc. Digital Audio Effects (DAFx).

[20] EKMAN P. 1993. Facial expression and emotion. American Psychologist 48: 384.

[21] GOLAN O., BARON-COHEN S. & HILL J. 2006. The cambridge mindreading (CAM) face-voice battery: testing complex emotion recognition in adults with and without Asperger syndrome. Journal of Autism and Developmental Disorders 36: 169-183.

[22] OERTEL C., SCHERER S. & CAMPBELL N. 2011. On the use of multimodal cues for the prediction of degrees of involvement in spontaneous conversation. Twelfth Annual Conference of the International Speech Communication Association.