# *Non-verbal Communication Training with an Interactive Multimedia Application*

Hiroki Tanaka, Sakriani Sakti, Graham Neubig, Tomoki Toda, Satoshi Nakamura

Nara Institute of Science and Technology, Japan

0188

The Asian Conference on Education 2013

Official Conference Proceedings 2013



iafor
The International Academic Forum
www.iafor.org

## 1. **Introduction**

Socialization and communication are important factors influencing human social life. People who have trouble with social skills and communication have recently been increasing due to environmental and/or inherent reasons [1]. Some papers mentioned the extreme case of these traits is autism [2], a set of neurodevelopmental conditions characterized by social interaction and communication difficulties, as well as unusually narrow, repetitive interests [3]. Given the impact of these problems on every- day life, there has been considerable interest in tools to both identify the degree of these difficulties and allow for training tool to improve social and communication skills [4].

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. It is well established that empathy, particularly as manifested in emotion recognition and mental state recognition, are core difficulties in people with Autism Spectrum Disorders (ASD).

There have been a number of studies on tools to train and test empathizing [5, 6]. In particular, we have proposed a tool a NOCOA (NOnver-bal COmmunication for Autism), which is communication aid application to help test and train the above skills [7]. NOCOA was designed according to several principles:

- 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 [8]." Fujisaki [9] uses the term "non-verbal" to refer to not only emotion, but also partner information, intention, situation, age, sex and other factors. Thus, the application should cover both emotion and other non-verbal behaviors.
- One of the factors influencing the ability to empathize is the severity of ASD. Autism is a spectrum condition [10] that has a broad range of clinical characteristics ranging from mild to severe. There are several methods such as the Autism Spectrum Quotient (AQ) for measuring a person's position on the autistic spectrum in both people with and without autism [11]. Thus, non-verbal behaviors as tested by NOCOA should have correlate with Autism Spectrum Quotient, and we have used this to guide our design.
- While, individuals with ASD have difficulty in socialization and communication, they also show good and sometimes even superior skills in "systemizing[2]." Systemizing is the drive to analyze or build systems, to understand and predict the behavior of events in terms of underlying rules and regularities. The use of systematic computer soft- ware for individuals with ASD can take advantages of this fact [12, 13]. Thus, NOCOA was designed to include a systematic training method.

However, while the overall design of NOCOA has proven advantages in our previous research, NOCOA used only speech data for testing and training non-verbal behavior. In contrast, previous research has found that communication difficulties have been found across different sensory modalities, both visual and auditory. Some reports mentioned both visual and audio information is important to recognize basic and complex emotion [14, 15]. In this paper, we expand upon the NOCOA framework, incorporating not only audio data, but also movie data. We include movies of the speaker. We refer to this updated application as NOCOA+.

## 2. NOCOA

In this section, we explain the design of NOCOA framework.

### 2.1. Assessment of Communication Skills and Socialization

Non-verbal behaviors include various factors (e.g., emotion, situation, and age). The objective of our previous work is to confirm the important non-verbal factors contributing to communication skills as measured by using AQ.
To do so, we used 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).

Next, we performed a factor analysis using individual responses to each question on the AQ questionnaire to determine several important factors for social and communication skills. Based on an analysis using principal component analysis (PCA) and the promax method we finally found 5 factors:
1. intention and interest
2. politeness or impoliteness as well as new friends.
3. social places and situations.
4. chit-chat and feelings.
5. other factors.

To confirm the degree to which each factor con- tributes to evaluating communication and social skills, we calculate Pearson's r value between each factor's total score. This revealed that the first five factors are sufficient to measure social and communication skills. Finally, we selected the first two factors (intention & interest, and politeness/impoliteness & new friends) as the non-verbal information to be trained and tested by NOCOA. These represent intention and partner information. The detail of categorization procedure is written in [7].

### 2.2 Structure

Once we had identified the important non-verbal factors that those with communication difficulties have trouble identifying we next had to design a systematic application to test and train those factors. To do so, we adopted a quiz format, where the user of the application must choose from several categories of intention and partner information.

There are two modes in NOCOA, Listening mode and Test mode. In listening mode, users touch the screen to choose the content, choose from two types of partner in- formation, and then choose from three types of intention. 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.

NOCOA also has a test mode, which is able to mea- sure users' intention and partner information cognitive skills. The user listens to the voice, and then chooses the appropriate intention 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 sets are chosen at random each time.

## 3. Movie Data

As mentioned in the previous section, NOCOA uses only audio data, but this is potentially limiting, in that both audio and visual information are important in real communication. NOCOA+ uses not only audio data but also movie data. In this section, the procedure of data collection and annotation are described.


Figure 1: Screenshot of NOCOA+.

### 3.1. Movie Recording

We designed our movie recording scenario so it would be easy to collect the non-verbal behaviors defined in the previous section in as natural a manner as possible. For intention, we would like to ensure that we can collect video samples of “derisive”, “social”, “friendly” speech. To do so we had each subject. Following procedure is implicated; 1) read the sports section of the newspaper, 2) converse about the content of the article for 10 minuets, 3) read the society section of the newspaper, 4) converse for 10 minutes. In addition, to make it easier to collect two types of partner information, we had each subject converse with both their good friend and their teacher.

In this study, 4 students (mean age 23.7) acted as subjects scored under 32 of AQ total (cut-off value of ASD). A video camera (SONY HDR-CX560) is used, and placed in the middle of the two conversants to take frontal shots. A pin microphone (Olympus ME52W) was used for recording a person’s speech data. Movie data and speech data are synchronized using Windows Movie Maker, and each speech interval (utterance) is detected using the power value extracted by Snack Tcl/Tk [19]. Detected utterances automatically divided into speech, movie, and audiovisual. We also created utterances including context information, the 5s and 10s prior to the actual utterance.

### 3.2. Classification

In movie recording, a total of 1200 audiovisual utterances are prepared. 3 annotators

who scored under 16 on the AQ test label the utterances through a web site. Sum of their AQ subareas scores for communication and social skills are each 1, 1, 1. Thus, the annotators likely have good social and communication skills. The annotators labeled each utterance into friend, teacher, or others for partner information and into derisive, social, friendly, or other for intention respectively. A total of 109 utterances were chosen for which all 3 annotators agreed for use in NOCOA+.

## 4. Design Modifications for NOCOA+

The iPad application NOCOA+ was developed towards social and communication skill support in real situations. Alike NOCOA, NOCOA+ has two modes, training mode and test mode. In this section, the main modifications to NOCOA implement in NOCOA+ are explained.

### 4.1. Training mode

Training mode is designed to enhance user's socialization and communication skills. Some papers speak of the extreme male brain theory of autism, which means that people with autism prefer a systematic experience [2]. A system is defined as something that takes input, which can then be operated on in a variety of ways, to deliver different outputs in a rule-governed way. Thus, we expanded training mode to provide two types of training, "listen to a large number of examples" and "check the rules". The former is developed to enable user to learn using a statistical-based training regimen, and the latter is a learning module using rules. The user can select the prefer one from the training menu.

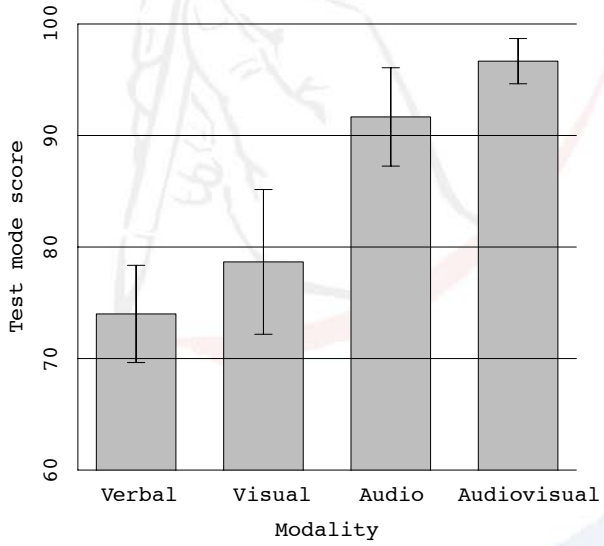


Figure 2: Score difference of each modality. Error bar shows standard error.

### 4.2. Test mode

In the test mode quiz, 10 questions for measuring user's non-verbal communication skills are provided. It has 2 types of generalization levels shown below:

- Closed: testing is performed using data that were included in the training mode
- Open: data is not included in the training mode

To show the previous annotation is valid, the three annotators take the test mode. The

annotators answer ten questions, which include four modalities, audiovisual, speech, movie, and verbal (where the first author of this paper transcribed the speech in the movie data and read it in a flat tone without emotion). The closed data was used, and scores were averaged. Figure 2 shows the result. The maximum score of test mode is 100, and in case of audiovisual, averaged score is above 95, indicating that the previous annotation is reliable. However, scores of speech, movie, and verbal are decrease compared to audiovisual. In both movie and verbal, it is significantly different comparison with audiovisual. In case of movie, relatively a large number of errors are found in politeness category, and in case of verbal, a large number of errors are found in intention category. Thus we can see that both audio and visual data are useful for identifying the non- verbal behaviors, and thus audiovisual is used for further investigations of test mode.

We further expand NOCOA by setting a difficulty level in test mode. We do this by having ten people other than the annotators use test mode. Three types of difficulty level is set; easy-normal-hard according to their accuracy rate. Accuracy rate of each difficulty level is as follows; easy: 81-100%, normal: 51-80%, hard: 0-50%. All utterances are classified into each difficulty level, and in the future if test mode results are sent to server, difficulty level is automatically updated according to accuracy rate. For further experiments, we use difficulty levels, easy and normal.

## 5. Experiment1: Measuring Non-verbal Communication Skills

### 5.1. Method

The purpose of Experiment 1 is to investigate the relationship between AQ score and non-verbal communication skills using NOCOA+ among members of the general population, and efficacy of movie data. First, eleven Japanese students (mean age: 23.3, 10 male and 1 female) enter a laboratory with little external stimulus, receive an explanation by the first author, perform each difficulty level: easy and normal in closed data one time, and the averaged score of two difficulty level is calculated. Finally, eleven students take Japanese version of AQ [20] (available at http://www.autism- communication.com/%7Ehiroki-tan/AQsub.html), and the sum of the two AQ subareas (communication and social skill) is measured.

### 5.2. Result

Figure 3 shows relationship between sum of social and communication scores and test mode score of NOCOA+. Maximum score of test mode is 100, and a high score indicates high non-verbal communication skills. The maxi- mum social and communication scores are each 10, and a high score indicates a high level of autistic traits. As Figure 3 shows, the correlation between sum of AQ subareas and averaged test mode score is high, with a correlation coefficient of 0.85 ($p$-value $< .01$). This reveals that large variations in the ability to recognize intention and partner information exist in the general population, and that these are significantly related to autistic traits. Note that despite the fact that the participants had not been diagnosed with Asperger syndrome or high-functioning autism, and have average or above average IQ, their range of AQ scores was wide and well correlated with test mode score. Be- cause in previous work with NOCOA, correlation coefficient was 0.71 ($p$-value $< .01$) [7], by incorporating movie data, the improvement of correlation is confirmed.

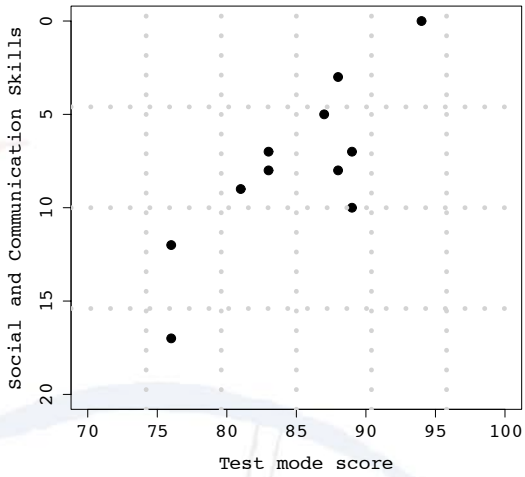


Figure 3: Relationship between sum of social communication score and test mode score of NOCOA+.

## 6. Experiment2: Effectiveness of Computer-Based Training

### 6.1. Method

The purpose of Experiment 2 is to investigate the efficacy of training mode among member of the general population. Maintaining high scores even in unseen open questions is also investigated. We recruited six students (mean age: 23, 5 male and 1 female). First, six students enter a laboratory, receive an description by first author, per- form each difficulty level: easy and normal in closed data one time. Then half of students who are randomly selected use training mode for 20 minutes, and the other half of students waited for the same 20 minutes. The training group is instructed to first use rule-based training and then use statistic-based training. Almost all students were able to complete all utterances in 20 minutes. After 20 minutes both the groups use test mode with both closed data and open data. The improved score between before and after 20 minutes of training or waiting is tested by Student's t-test.

### 6.2. Result

Figure 4.2 shows the improvement of test mode score before 20 minutes and after 20 minutes. In terms of difficulty level easy (left side of Figure 4.2), the improvement in score is 7.66 in training group and -2.33 in non-training group respectively. Improvement in score on open data, which means score between before 20 minutes on training set and after 20 minutes on test set, is 9.66 in the training group and -1.33 in the non-training group respectively. In terms of difficulty level normal (right side of Figure 4.2), the improvement in score 12.00 in training group and -3.00 in the non-training group respectively. The improvement on open data is 7.33 in the training group and 5.66 in non-training group respectively.

The result shows that in terms of difficulty level easy, 20 minutes training is effective with both closed data and open data, and in both audio data and movie data, we confirm improvement of test mode score by systematic training. However, in terms of difficulty level normal, questions not include in training mode is still difficult. For this difficulty, because people with autism have difficulties in generalizing learned social skills [23, 24, 25], autistic traits which individuals have should be considered.

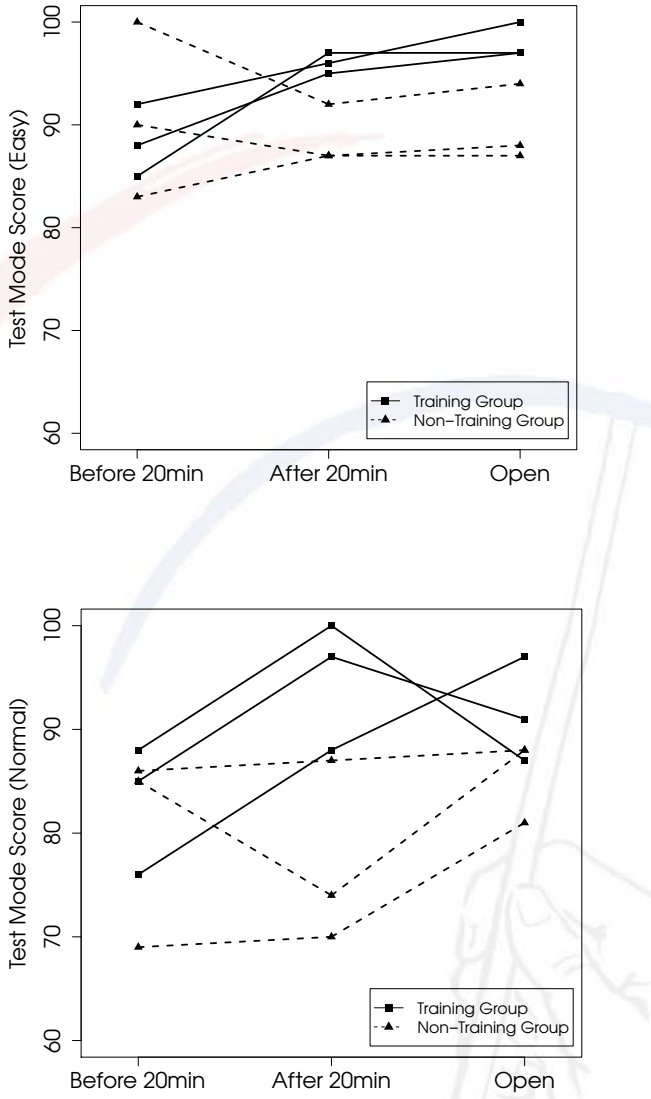


Figure 4: Left figure indicates difficulty level easy, and right figure indicates difficulty level normal. Dot-line is score of non-training group, and solid-line is score of training group. Pre and Post 20 minutes (Closed data) is shown as well as Post 20 minutes (Open data).

## 7. Conclusion

Previous research has found that social and communication difficulties have been found across different sensory modalities, both visual and auditory. We expand upon the NOCOA framework with not only audio data, but also movie data. We refer to this updated application as NOCOA+. In this paper we confirm the relationship between non-verbal communication skills and AQ subareas by using NOCOA+, and examine prospective intervention through teaching non-verbal communication skills, intention and partner information.

One potential direction for the future is considering individual differences (e.g. relationship between tendency of mistakes and autistic traits). Collecting more data

for training is also important, and as a substitute plan, contextual data can be used for training. In addition, people with ASD need long term treatment for emotion recognition [17]. Thus we will design the next experiment with long-term follow-up in people with ASD.

## 8. Acknowledgement

We thank Dr. Hidemi Iwasaka at Nara University of Education, for his advice and support.

## 9. References

[1] GOLEMAN D., 2007. Social intelligence. Arrow Books.
[2] BARON-COHEN S., RICHLER J., BISARYA D., GURUNATHAN N., WHEELWRIGHT S., BARON-COHEN S., RICHLER J., BISARYA D., GURUNATHAN N. & WHEELWRIGHT S. 2003. The systemizing quotient: an investigation of adults with Asperger syndrome or high―functioning autism, and normal sex differences. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 358: 361-374.
[3] KANNER L., 1943. Autistic disturbances of affective contact. Nervous Child 2: 217-250.
[4] BARON-COHEN S., 2008. Autism and Asperger syndrome. Oxford University Press, USA.
[5] BAUMINGER N., 2002. The facilitation of social-emotional understanding and social interaction in high-functioning children with autism: Intervention outcomes. Journal of autism and developmental dis- orders 32: 283-298.
[6] GOLAN O., & BARON-COHEN S., 2006. Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. Develop. Psychopatholgy.
[7] TANAKA H., SAKTI S., NEUBIG G., TODA T., & NAKAMURA S., 2012. Non-verbal cognitive skills and autistic conditions: An analysis and training tool, In Proc IEEE CogInfoCom.
[8] AMERICAN PSYCHIATRIC ASSOCIATION, 1994. The diagnostic and statistical manual of mental disorders, IV . Washington, D.C.: American Psychiatric Association.
[9] FUJISAKI H., 1997. Prosody, models, and spontaneous speech. Computing Prosody.
[10] WING L., 1996. Autistic spectrum disorders. Bmj 312: 327.
[11] BARON-COHEN S., WHEELWRIGHT S., SKIN- NER R., MARTIN J., & CLUBLEY E., 2001. The Autism-Spectrum Quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. Journal of Autism and Developmental Disorders 31: 5-17.
[12] BISHOP J., 2003. The Internet for educating individuals with social impairments. Journal of Computer Assisted Learning 19: 546-556.
[13] 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.
[14] GOLAN O., BARON-COHEN S., HILL J. J., & RUTHERFORD M. D., 2007. The‘ reading the mind in the voice ’test-revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. Journal of autism and develop- mental disorders 37: 1096-1106.
[15] GOLAN O., BARON-COHEN S., & GOLAN Y., 2008. The‘ Reading the Mind in Films ’task [child version]: Complex emotion and mental state recognition in children with and without autism spectrum conditions. Journal of Autism and Developmental Disorders 38: 1534-1541.
[18] TANAKA Y., & WAKIMOTO K., 1983. Methods of multivariate statistical analysis. Gendai-Sugoku, Tokyo.
[19] The Snack Sound Toolkit. http://www.speech.kth.se/snack/
[20] WAKABAYASHI A., BARON-COHEN S., WHEELWRIGHT S., & TOJO Y., 2006. The Autism-Spectrum Quotient (AQ) in Japan: a cross-cultural comparison.

Journal of Autism and Developmental Disorders 36: 263-270.
[21] EKMAN P., 1993. Facial expression and emotion. American Psychologist 48: 384.
[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.
[23] BLTE S., FEINEIS-MATTHEWS S., LEBER S., DIERKS T., HUBL D., & POUSTKA F., 2002. The development and evaluation of a computer-based program to test and to teach the recognition of facial affect. International Journal of Circumpolar Health 61:
[24] SILVER M., & OAKES P., 2001. Evaluation of a new computer intervention to teach people with autism or Asperger syndrome to recognize and predict emotions in others. Autism 5: 299-316.
[25] BERNARD-OPITZ V., SRIRAM N., & NAKHODA-SAPUAN S., 2001. Enhancing social problem solving in children with autism and normal children through computer-assisted instruction. Journal of Autism and Developmental Disorders 31: 377-384.

iafor