INFORMATION AND SYSTEMS IN EDUCATION, Vol. 12, No. 1 2013
pp. 19–26

NOCOA: A Computer-Based Training Tool for Social and Communication Skills That Exploits Non-verbal Behaviors

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

(Received 12 April 2013 and accepted in revised form 17 October 2013)

Abstract The number of people with social skills and communication difficulties is now greater than ever for a variety of reasons. Our objective is to measure these difficulties, and enable people with these difficulties to improve their social and communication skills for use in the real world. This paper examines the relationship between non-verbal communication skills and the autism spectrum quotient among members of the general population. We also propose a training framework for these skills. 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 communication skills for members of the general population.

***Keywords*:** non-verbal communication, socialization, computer-based training, autism-spectrum quotient

1. Introduction

The numbers of people with social skills and communication difficulties have recently been increasing for various reasons(1). These difficulties cause problems in relating to other people, and are central characteristics of autism(2). Technology for both identifying the degree of these difficulties and developing a learning tool for social and communication skills could help people overcome these barriers.

Autism is a set of neurodevelopmental conditions characterized by social interaction and communication difficulties, as well as unusually narrow and repetitive interests(3). The diagnosis criteria of autism(4) 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." Fujisaki(5) uses the term non-verbal to refer to not only emotion, but also eye contact, intention, gesture, gender and other factors.

One of the psychological themes in autism is empathizing. Empathizing is a set of cognitive and affective skills people use to make sense of and navigate the social world. The cognitive component of empathy is also referred to as the theory of the mind(6). It is well established that emotion recognition and mental state recognition are core difficulties in people with Autism Spectrum Disorders (ASD). Such difficulties have been found across different sensory modalities, both visual and auditory(7).

Neuroimaging studies of emotion recognition from faces reveal that people with ASD show less activation in brain regions central to face processing, such as the fusiform gyrus(8). 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 ASD process socioemotional information(8, 9).

One of the factors influencing the ability to empathize is the severity of ASD. Autism is a spectrum condition(10); that means it 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)(11) is a self-administered screening measurement that can be used for children from four years of age through to adulthood. It has a total of 50 statements; that is, 10 questions each assessing 5 different areas: social skills; attention switching; attention to detail; communication; and imagination. In the AQ, one statement scores one point if the respondent records abnormal or autistic-like behavior either mildly or strongly. Thus individuals score in the range of 0–50. However, a high AQ score alone is not a reason to seek a professional diagnosis. 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 ASD. Among members of the

* Graduate School of Information Science, Nara Institute of Science and Technology, Japan

general population, autistic conditions are widely distributed across a spectrum.

In contrast to these difficulties, individuals with ASD show good and sometimes 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. Learning empathizing is important for social skills training[12, 13]. However, training programs for this typically do not focus specifically on systematically teaching emotion, but instead address other issues, such as conversation, reducing socially inappropriate behavior, and so forth. The use of computer software for individuals with ASD has several advantages: first, individuals with ASD favor the computerized environment because it is predictable, consistent, and free from social demands, which they may find stressful. Second, users can work at their own pace and level of understanding. Third, lessons can be repeated over and over again, until mastery is achieved. Fourth, interest and motivation can be maintained through different and individually selected computerized rewards[14, 15].

Golan and Baron-Cohen[16] suggested the Mindreading DVD, which enables adults with autism to systematically learn mental state recognition and they found an improvement in emotion recognition skills was achieved through several months training. However other generalization levels (questions not include in training) were still difficult, and these typically do not include non-verbal signals.

Taking into consideration the issues mentioned above, we attempt to first clarify the relationship non-verbal communication skills and autistic conditions and then to develop a systematic training method for social and communication skills. For the first goal, we evaluate adult AQ scores to confirm the non-verbal factors contributing to social and communication skills, which include, but are not restricted to, emotion. For the second goal, we develop a mobile application reflecting the result of this analysis of AQ tendencies. The mobile application allows users to measure autistic traits automatically, and enables people with social and communication difficulties to improve non-verbal communication skills for use in the real world.

2. Assessment of Communication Skills

Non-verbal information includes various factors (e.g., eye contact, intention, gesture, and gender). The objective of this section is to confirm the important non-verbal factors contributing to communication skills as measured by AQ. To collect AQ data, we first recruited 21 Japanese students to take the English version of the AQ. We expected that three AQ areas, attention switching, attention to detail, and imagination, are not related to the communication difficulties. Thus we measured the other two of the original five areas: social and communication skills (with a total of 20 statements). Each question is listed in Table 1.

First, we defined the factor class number based on principal component analysis and the chi-square value which resulted in five factor classes[18]. After that we performed a factor analysis with maximum likelihood estimation to determine several important factors in each class for social skills and communication based on the AQ[19]. Table 1 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 performed an analysis with the promax method, which is an alternative non-orthogonal (oblique) rotation method that is effective when there are highly correlated factors. This revealed the following points.

- The first factor is largely related to intention and interest.
- The second is related to politeness or impoliteness as well as new friends.
- The third is related to social places and situations.
- The fourth is related to chit-chat and feelings.
- The fifth is other factors.

As a result we selected the first two factors (intention, interest, and politeness/impoliteness, new friends) as non-verbal information, and named these groupings as representing "intention" and "partner information", respectively.

3. Classification of Natural Speech Data

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 communication skills.

3.1 Natural conversational speech corpus

The FAN subset of JST/CREST Expressive Speech

Table 1. Factor Fnalysis Using the Promax Rotation Method[17]. Columns One to Five Show the Loadings and the Proportion of Variance from the First to the Fifth Factors. The Left-most Column Indicates the Original AQ Question Number of Each 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 …		–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 …	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

Processing (ESP) corpus (www.speech-data.jp) 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 onto 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[20]. We prepared a total of 5,367 short utterances from the FAN database following previous work.

3.2 Communication skill categorization

Based on the results of the factor analysis, we decided 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 three types of categories by three Japanese male students (age: 23–24). The final total number of contents of the utterances was 27. The final categories of partner information were “friend” and “teacher”, and categories for intention were “derisive”, “social”, and “friendly”. The categorization procedure was as follows.

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

4. Mobile Application NOCOA

We developed an iPad application named NOCOA(21), which reflects the above result of overall AQ tendencies and classification of short utterances.

4.1 Facial images

We prepared facial images for each category. As we described, we chose three axes: 27 types of contents of the utterances, three types of intentions or interests of talk, and two types of partner information. Both actual human photos (chosen via yourstock.com) and illustrations were prepared for each category, and the use of photos or illustrations can be chosen by the user.

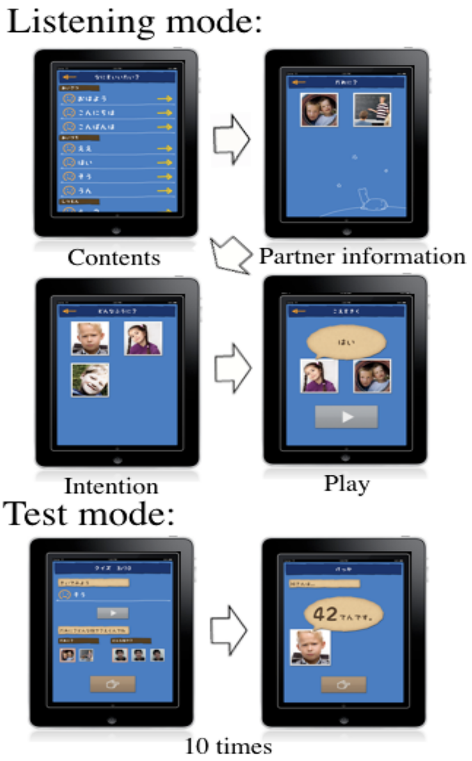


Figure 1. Two Modes of NOCOA, Listening Mode and Test Mode.

4.2 Structure

This subsection explains two modes of NOCOA (Figure 1). Both modes were developed systematically.

4.2.1 Listening mode

In listening mode, users touch the screen to choose the content (frequently-appearing utterances from the corpus), choose from two types of partner information, and then choose from three types of intention (Figure 2). After selection of the content, partner information, and intention, the play screen will be shown. The user can see the result he/she chose on the play screen, and can listen to the appropriate sound. The maximum number of sounds in each category is four, and the sound is

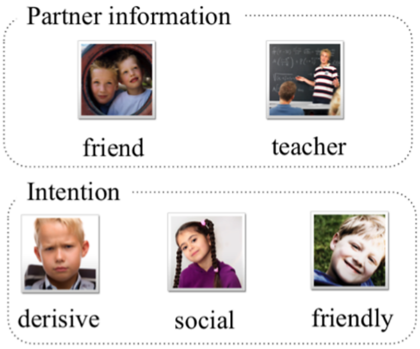


Figure 2. The Images Used in NOCOA.

played randomly.

4.2.2 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 five. For partner information, the user gets a score of five when the correct partner is chosen and zero otherwise. For intention, the score for mistakes between derisive and social is two, between social and friendly is three, and between derisive and friendly is zero. The test mode score is calculated after answering ten questions, so 100 is the best score. The ten question set is chosen at random each time. Test mode also has three generalization levels:

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

5. Experiment and Evaluation

5.1 Experimental setting

We did an experiment to get the correlation between AQ score and test mode score in members of the general population. We did this because our tool was developed for people who have difficulties with social and communication skills to measure their non-verbal communication skills and to systematically learn how to identify non-verbal information. The procedure of the experiment was as follows: 19 Japanese participants were recruited (18 males and 1 female; mean age: 25.0). 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. We confirmed that all participants did not have difficulty in understanding the concepts. Then, participants took the generalization level 1 (closed data) of the test mode on NOCOA two times, and the average score of two trials was calculated.

Here, computer-based intervention used drawings of photographs for training, rather than more lifelike stimuli. This might have made generalization harder than if more ecologically valid stimuli were used. We also tested efficacy of listening mode with several Japanese students (training group) who scored below the average (all males; mean age: 23.0). They used the listening mode for 20 minutes, and the control group waited for the same 20 minutes. After 20 minutes both groups used the test mode with the three generalization levels.

5.2 Experimental results

First, we measured the relationship between the test mode score and AQ. Figure 3 shows that the correlation coefficient between AQ and averaged test mode score was 0.71 ($p<.01$). This means that large variations in the ability to recognize non-verbal and partner information exist in the general population, and that is significantly related to evaluation of autistic traits. We note that despite the fact that the participants had not been diagnosed with Asperger syndrome or high-functioning autism their range of AQ scores was wide and well correlated with test mode score.

We also confirmed efficacy of listening mode with two Japanese students who scored below average (mean age: 23.0) and participated in training. Figure 4 shows that after using the listening mode for 20 minutes, their score also improved above ten points in the case of generalization level 1. As a result of training we found they maintained a high score in both open and long utterances (Figure 5).

6. Conclusion

In this paper, we confirmed the relationship between non-verbal communication skills and AQ by using speech output with visual hints and examined prospective intervention through systematically teaching non-verbal information, intention and partner information. According to a factor analysis, we confirmed two important axes. Previous reports mention basic or complex emotions but not partner information(7, 16, 22). We conducted a subjective experiment with members of the general population. From the experiment, we determined the correlation between AQ score and test mode score

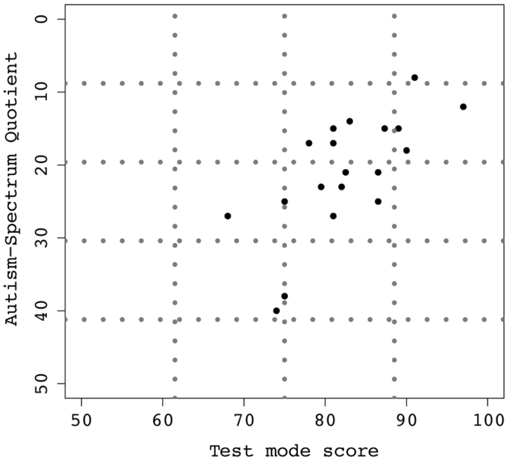


Figure 3. Relationship between Test Mode Score and AQ.

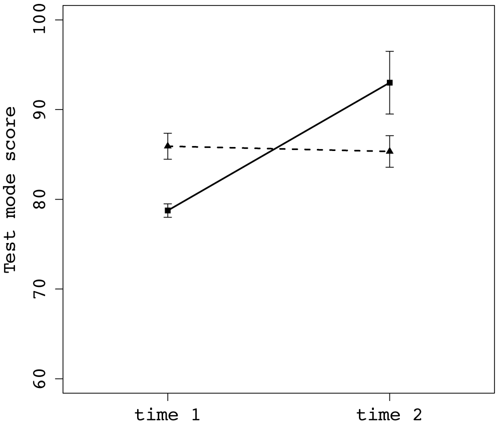


Figure 4. The Test Mode Scores before (time 1) and after (time 2) the Training with Standard Error Bars. The Dotted Line Shows the Control Group and the Solid Line Shows the Training Group.

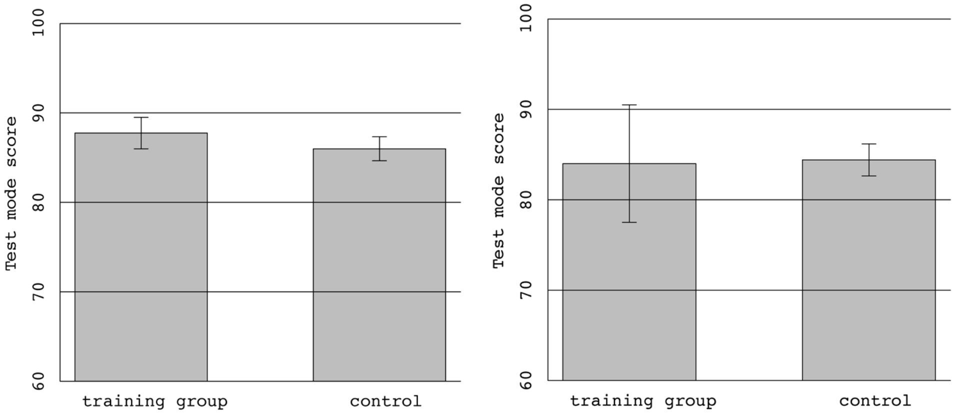


Figure 5. The Scores of the Training Group and Control Group. The Left Graph Shows the Result in Generalization Level Two and the Right One Shows the Result in Generalization Level Three.

was 0.71 for 19 Japanese adults. This shows that ASD severity is significantly related to test mode score even in the Japanese adult group studied. It also reveals that in the general population, where the range of AQ scores is wider, that the more autistic traits a person possesses, the more difficult it becomes to recognize non-verbal information. In addition several Japanese student participants had difficulty distinguishing utterances compared to other participants. However their test mode score was improved by using the listening mode for 20 minutes.

Our goal is to further improve the training tool towards supporting real communication. Though this paper presented visual hints as static graphical images, we will try to use movie data in the future.

Acknowledgement

We thank Dr. Hidemi Iwasaka at the Research Center for Special Needs Education, Nara University of Education, for his advice and support.

References

(1) Goleman, D.: Social Intelligence: The New Science of Human Relationships, Arrow Books, London (2007).

(2) Baron-Cohen, S.: Autism and Asperger Syndrome, Oxford University Press, New York (2008).

(3) Kanner, L.: “Autistic Disturbances of Affective Contact”, Nervous Child, Vol. 2, pp. 217-250 (1943).

(4) American Psychiatric Association: The Diagnostic and Statistical Manual of Mental Disorders, IV. American Psychiatric Association, Washington, D.C. (1994).

(5) Fujisaki, H.: “Prosody, Models, and Spontaneous Speech”, in Computing Prosody, eds. Sagisaka, Y., Campbell, N. and Higuchi, N., pp. 27–42. Springer, New York (1997).

(6) Astington, J. W., Harris, P. L. and Olson, D. R. (eds.): Developing Theories of Mind, Cambridge University Press, Cambridge (1988).

(7) Golan, O., Baron-Cohen, S. and Hill, J.: “The Cambridge Mindreading (CAM) Face-voice Battery: Testing Complex Emotion Recognition in Adults with and without Asperger Syndrome”, J. of Autism and Developmental Disorders, Vol. 36, pp. 169–183 (2006).

(8) Critchley, D., Daly, M., Bullmore, T. et al.: “The Functional Neuroanatomy of Social Behaviour”, Brain, Vol. 123, pp. 2203–2212 (2000).

(9) Ashwin, C., Baron-Cohen, S., Wheelwright, S. et al.: “Differential Activation of the Amygdala and the Social Brain during Fearful Face-processing in Asperger Syndrome”, Neuropsychologia, Vol. 45, pp. 2–14 (2007).

(10) Wing, L.: “Autistic Spectrum Disorders”, British Medical J., Vol. 312, pp. 327–328 (1996).

(11) Baron-Cohen, S., Wheelwright, S., Skinner, R. et al.: “The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/high-functioning Autism, Males and Females, Scientists and Mathematicians”, J. of Autism and Developmental Disorders, Vol. 31, pp. 5–17 (2001).

(12) Bauminger, N.: “The Facilitation of Social-emotional Understanding and Social Interaction in High-functioning Children with Autism: Intervention Outcomes”, J. of Autism and Developmental Disorders, Vol. 32, pp. 283–298 (2002).

(13) Ozonoff, S. and Miller, N.: “Teaching Theory of Mind: A new approach to social skills training for individuals with autism”, J. of Autism and Developmental Disorders, Vol. 25, pp. 415–433 (1995).

(14) Bishop, J.: “The Internet for Educating Individuals with Social Impairments”, J. of Computer Assisted Learning, Vol. 19, pp. 546–556 (2003).

(15) Moore, D., McGrath, P. and Thorpe, J.: “Computer-aided Learning for People with Autism—A Framework for Research and Development”, Innovations in Education and Teaching International, Vol. 37, pp. 218–228 (2000).

(16) Golan, O. and Baron-Cohen, S.: “Systemizing Empathy: Teaching Adults with Asperger Syndrome or High-functioning Autism to Recognize Complex Emotions Using Interactive Multimedia”, Development and Psychopathology, Vol. 18 (2006).

(17) Tanaka, Y. and Wakimoto, K.: Methods of Multivariate Statistical Analysis, Gendai Sugoku, Tokyo (1983).

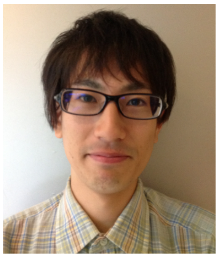
(18) Yanagii, H. and Takagi, H.: Handbook of Actual Examples of Multivariable Analysis, Gendai-Sugakusha, Tokyo (1986) (in Japanese).

(19) Oshio, S.: The Analysis of Psychological and Survey Data by SPSS and Amos, Tosho, Tokyo (2004).

(20) Campbell, N.: “Conversational Speech Synthesis and the Need for Some Laughter”, IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, pp. 1171–1178 (2006).

(21) Tanaka, H., Sakti, S., Neubig, G. et al.: “Non-verbal Cognitive Skills and Autistic Conditions: An Analysis and Training Tool.” In Proc. 3rd IEEE CogInfoCom. pp. 41–46 (2012).

(22) Ekman, P.: “Facial Expression and Emotion”, American Psychologist, Vol. 48, p.384 (1993).



Hiroki Tanaka received his B.E. from Asahikawa National College of Technology in 2010, and his M.E from Nara Institute of Science and Technology in 2012. He is currently Ph.D. student at Nara Institute of Science and Technology, and researcher at the Research Center for Special Needs Education, Nara University of Education. His research interests include education systems for non-verbal communication, and automatic measurement of communication skill.

Sakriani Sakti received her B.E degree in Informatics from Bandung Institute of Technology, Indonesia, in 1999. She received her MSc degree from University of Ulm, Germany, in 2002. She continued her study (2005–2008) with Dialog Systems Group University of Ulm, Germany, and received her PhD degree in 2008. Currently, she is an assistant professor of the Augmented Human Communication Lab, NAIST, Japan. Her research interests include statistical pattern recognition, speech recognition, spoken language translation, cognitive communication, and graphical modeling framework.

Graham Neubig received his B.E. from University of Illinois, Urbana-Champaign, U.S.A, in 2005, and his M.E. and Ph.D. in informatics from Kyoto University, Kyoto, Japan in 2010 and 2012 respectively. He is currently an assistant professor at the Nara Institute of Science an Technology, Nara, Japan. His research interests include speech and natural language processing, with a focus on machine learning approaches for applications such as machine translation, speech recognition, and spoken dialog.

Tomoki Toda earned his B.E. degree from Nagoya University, Aichi, Japan, in 1999 and his M.E. and D.E. degrees from the Graduate School of Information Science, NAIST, Nara, Japan, in 2001 and 2003, respectively. He was an Assistant Professor of the Graduate School of Information Science, NAIST from 2005 to 2011, where he is currently an Associate Professor. His research interests include speech and language processing, such as speech synthesis. He received 11 awards including the IEEE SPS 2009 Young Author Best Paper Award and the EURASIP-ISCA Best Paper Award 2013.

Satoshi Nakamura is Professor of Graduate School of Information Science, Nara Institute of Science and Technology, Japan, Honorar professor of Karlsruhe Institute of Technology, Germany, and ATR Fellow. He received his B.S. from Kyoto Institute of Technology in 1981 and Ph.D. from Kyoto University in 1992. He organized the International Workshop of Spoken Language Translation (IWSLT 2006) and Oriental Cocosda 2008 as a general chair. He also served as the program chair of INTERSPEECH 2010. He has been Elected Board Member of International Speech Communication Association, ISCA, since June 2011 and IEEE Signal Processing Magazine Editorial Board Member since April 2012.