Mobile Tutoring System in Facial Expression Perception and Production for Children with Autism Spectrum Disorder

Sergey Anishchenko[1], Alexandra Sarelaynen[1], Konstantin Kalinin[1], Anastasia Popova[1], Natalia Malygina-Lastovka[2] and Kira Mesnyankina[2]

[1]*Southern Federal University, Stachki ave., Rostov-on-Don, Russia*

[2]*Assistance Centre for Families of Children with Special Needs "Sodejstvie", Rostov-on-Don, Russia*

sergey.anishenko@gmail.com, {aisarelaynen, kkalinin, anastasp}@sfedu.ru, lastovkanata@mail.ru, k.mesnyankina@yandex.ru

Keywords: Facial Expression Recognition, Action Unit Recognition, Autism.

Abstract: Children with autism spectrum disorder are impaired in their ability to produce and recognize facial expressions and unable to interpret the social meaning of facial cues. Human interventionists can effectively train autistic individuals on facial expressions perception and production, they may benefit even more from computer-based intervention. In this study the tablet PC application was developed for learning facial expression perception and production. It uses newly designed computer vision algorithm for facial expression analysis which allows to estimate if the posed expression correct or not and guide users. Clinical trial was done with participation of 19 volunteer subjects from 6 to 12 years old. It was shown that after intervention subjects skills in emotion recognition were improved. Ability to transfer newly developed skills to children's everyday life was also investigated. Parent's questioning performed in 6 months after the intervention demonstrates that 10 out of 19 children were able to recognize emotion and respectively change their behavior in everyday life.

1 INTRODUCTION

Recent advances in facial expression analysis with computer vision tools open huge opportunities for investigation of issues that have previously been intractable in the field of behavioural research.

Computer vision technologies become popular in the investigation of facial expression recognition and production deficits common to children with autism spectrum disorder (ASD) (Cockburn et al., 2008; Gordon, Pierce, Bartlett, & Tanaka, 2014; Hashemi et al., 2014; Jain, Tamersoy, Zhang, Aggarwal, & Orvalho, 2012). Besides, low-cost systems which are under development now aim to replace high level specialist (Hashemi et al., 2014) in early autism diagnosis.

ASD is clinically diagnosed as impaired socialization and communicative abilities in the presence of restricted patterns of behavior and interests. Children with ASD are impaired in their ability to produce and perceive dynamic facial expressions (Adolphs, Sears, & Piven, 2001) and unable to interpret the social meaning of facial cues.

Several studies have already demonstrated that human interventionist can effectively train individuals with an ASD on facial expressions perception and production (DeQuinzio, Townsend, Sturmey, & Poulson, 2007). Although efficient, these methods are time-intensive and require high level specialist, thus they are burdensome.

Children with ASD may benefit more from computer-based intervention than from conventional methods which rely solely on high level specialist (Grynszpan, Weiss, Perez-Diaz, & Gal, 2014; Heimann, Nelson, Tjus, & Gillberg, 1995). In object naming skills development it was found that children following computer-based instruction learned significantly more new words and showed greater motivation for learning than children who experienced conventional approach (Moore & Calvert, 2000).

There are two groups of approaches in computer vision for facial expression analisys. First relies on geometric position of facial points, second on the face appearance. Most of the algorithms are designed to recognize unknown expression or Action Unit (AU) state shown in the photo. For a

Anishchenko S., Sarelaynen A., Kalinin K., Popova A., Malygina-Lastovka N. and Mesnyankina K.
Mobile Tutoring System in Facial Expression Perception and Production for Children with Autism Spectrum Disorder.
DOI: 10.5220/0006146003190324
In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, pages 319-324
ISBN: 978-989-758-226-4

training tool the task of computer vision algorithm is to estimate if the expression given in the photo matches to particular known expression. For ability of guiding user in expression production the algorithm also has to estimate what AU is incorrect if any and in what way. Besides, the intensity of expression also has to be estimated to make sure it is in some average range and not atypically intensive.

Appearance-based method can provide some information on the emotion intensity. Intensity correlates with distance from the input feature vector to the hyper plane learned by SVM (Littlewort et al., 2011). Taking into account that the photo with neutral face is available for training tool, more natural way to estimate intensity is to analyze facial point geometric displacement distance.

The aim of this work is to (1) develop and investigate low-cost computer vision tools for tablet PC which allows to measure correctness of facial expression and each AU separately and guide user to product correct expression; (2) investigate influence of the training tool on childrens ability to recognize facial expressions; (3) estimate the difference in computer-based and conventional learning of expression perception and production; (4) investigate how the emotion recognition skills transfer into children's everyday life.

This work examines potential benefits that computer vision along with low-cost mobile devices can provide for teaching children with ASD to percept and produce facial expressions. It is a first milestone in a project on developing reliable tools for teaching autistic children socialization without high level specialist.

2 METHODOLOGY

The developed training tool is the tablet PC (iPad>=2 with iOS>=8, Apple Inc.) application. It consists of two parts. First is the set of images for learning significant facial areas for six basic emotions (happiness, surprise, sadness, fear, disgust, anger) (Ekman & Rosenberg, 2012). Second is the tool for training emotion production by imitation. In the second part the photo with sample is displayed on the screen and user has to pose the same expression then press the button for analysis. Computer vision algorithm performs detailed analysis of each AU and guides the user in case of mistake by highlighting incorrect AU in his photo with animated contour.

2.1 Facial Expression Assessing

The overall idea of the face expression analysis in the developed tool is to estimate AUs state comparing geometric distribution of facial feature points in the input photo with one in the neutral photo and after that check if the AUs state is within predefined range for the particular emotion which child choose for training.

Face contour is detected by means of the Active Shape Model implemented in CSIRO Face Analysis SDK (M. Cox, J. Nuevo, 2013). Once the contour detected it is normalized by scaling. Two points which location is least affected by expression are used for computing the scaling coefficient (Fig. 1). All facial images are normalized so that the distance between the abovementioned points is 250 px. We assume that face is in near frontal orientation hence no head pose adjustment performed.

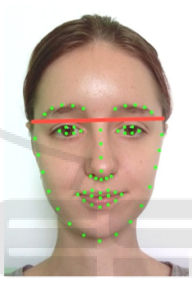

Figure 1: The dots show the face shape detected by CSIRO, while the line demonstrates the points used for computing coefficient for normalization by scaling.

After normalization six parameters (one per AU) are computed from the face shape (Table 1). The difference between corresponding AUs parameters in neutral and input images is used as characteristic of the state of AUs. If the state of the AU is within the predefined range for particular emotion then it is considered as correct one, otherwise AU is incorrect. If all AUs corresponding to particular emotion are correct then expression is also correct.

Correspondence of AUs and emotion in this project is in line with (Wallace & Ekman, 1983). The predefined range for each AU in each emotion was computed on the basis of the Warsaw Set of Emotional Facial Expression Pictures (Olszanowski et al., 2015).

Along with correctness estimation the developed approach allows to figure out how particular AU must be changed to match the emotion. It enables possibility to guide child during training. This is the main reason why this approach was chosen instead of classifier-based one which allows only binary classification.

Table 1: List of parameters.

#	Parameter	Description
1	Eye openness	The sum of the distances between the opposite points on upper and lower eyelids (Fig. 2 top left).
2	Eyebrow raised	The sum of the distances between the point on a nose bridge and five points on the eyebrow contour (Fig. 2 top right).
3	Nose furrow	The distances between the point on the nose basement and points on the nostril wings (Fig. 2 bottom left).
4	Mouth stretch	The distance between the rightmost point and the leftmost point of the mouth.
5	Upper lip raised	The distance between the point on the nose basement and the middle point of the upper lip outer contour (Fig. 2 bottom right).
6	Mouth openness	The sum of the distances between five opposite points on the outer lip contour.

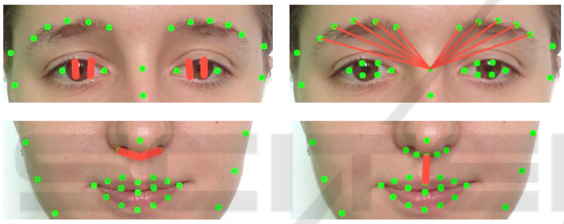

Figure 2: The scheme of the AU parameters computed from the face shape.

Photo of the neutral face of each child is taken and AUs parameters are extracted just once before training. The developed method of AU estimation does not need any image set for training classifier because it is based on comparison of neutral and input face shape.

There are some limitations in the face shape fitting algorithm which do not allow to estimate the state of AU in some expressions. In particular the model of the mouth shape variation does not cover corners lowering. Hence, the emotion of sadness cannot be assessed without extra algorithm for more detailed analysis of the AU. Therefore, guiding for production of sadness was solely relying on the trained specialist.

2.2 Training Tools

The described above computer vision algorithm is implemented in C++ and integrated into the tablet PC application which is used in the following way.

Before training a child takes a picture with the neutral face which is used for AU state analysis. At the beginning of training he has to choose emotion to train. The training itself consists of learning expressions and their production.

Learning part consists of the emotional face images specially designed to emphasize areas with important AU (Fig. 3 left, center). The child has to swipe and examine all images with highlighted AUs important for the selected expression. At the end the child has to pass a little test by selecting emoticon corresponding to the expression (Fig. 3 right).

Figure 3: Facial images for learning expression (happiness). *Left and centre* pictures contain a bounding box for which is highlighting the significant areas (mouth and eyes). The area besides the box is blurred. *Right*: emotional face with emoticons.

In the production part of the tool a photo with a sample of the emotion is demonstrated to the child and at the same time he sees himself in the viewfinder of the frontal camera (Fig. 4). He has to pose emotion and take a picture of himself by pressing button. The photo is processed as described in sec. 2.1 and if the expression is incorrect the user is guided by animated AU contour.

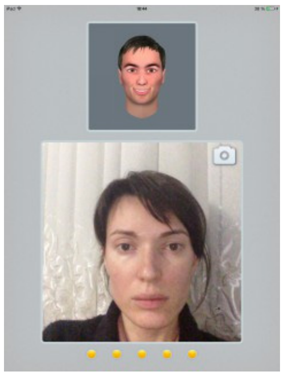

Figure 4: The screen in the tutoring application for learning of expression production. It contains the image with the sample expression (top) and camera viewfinder (bottom).

Data for training as well as information on results are stored on the remote server. It has a web interface for data managing and logs analyzing. The application can be synchronized with the server using API accessed over HTTP.

2.3 Data Collection and Analysis

The study was done with autistic children at the Assistance Centre for Families of Children with Special Needs "Sodejstvie", Rostov-on-Don, Russia by the licensed specialists. Approval for this study was obtained from the Institutional Review Board at the Southern Federal University, Rostov-on-Don, Russia. Parents and children were instructed and their signed agreements were collected before the study.

Nineteen volunteer subjects participated in the study, 2 girls and 17 boys, from 6 to 12 years old. Before the study children underwent neuropsychological and psychological tests with the purpose to estimate their ability to participate in the study. The trained specialist observed the children during the study constantly. He gave instructions before training and controlled the overall process.

Advantage of the tablet PC is in its wide availability and mobility. Besides, it motivates children to be involved in the training. The drawback is unconstraint illumination and poses that increase the error rate in the image processing algorithms.

Trained specialist performed constant observation of the training process and in case of software failure gave respective instructions to prevent children's frustration. Beforehand children were instructed that the trainer's instructions are more important than the tablet PC ones.

Subjects were separated into two groups for which different type of intervention was applied. The *first group* children (*n=13*) participated in the training with the tablet PC (below referred to as PC-based intervention). The *second group* (*n=6*) participated in the study designed for the same purpose (learn expression perception and production) but without any devices (below referred to as conventional intervention). The printed photo and mirrors were used to train children in the second group. The purpose of subject separation is to estimate the difference in PC-based and conventional interventions.

Before and after both kinds of interventions the children passed the following test to estimate and quantify their skills. They examined 24 printed photos and sketches (4 pictures per one basic emotion), which they had never seen before, with the emotional faces and tried to recognize emotion. Hence the study consisted of 3 steps. The *first step* is designed for testing initial ability of emotion recognition. The *second step* is the training either with or without tablet PC. The *last step* is designed to test the ability to recognize expression after the intervention. Besides, in 6 months after the interventions parents were questioned to assess children's ability to transfer their skills to everyday life.

3 RESULTS AND DISCUSSION

The difference in children's skills in emotion recognition before and after PC-based intervention was estimated by computing average number of correct answers in test with 24 photos. The result was 16.62 (4.11) and 22.46 (2.93) respectively (value in parentheses indicates standard deviation). Obviously after the intervention the skills in emotion recognition were improved. Analysis of correct answers separately for each emotion reviled that initially children could recognize happiness and sadness quite well (Fig. 5, emotion #1 and 2). After intervention they could recognize all emotions at similar level (Fig. 5).

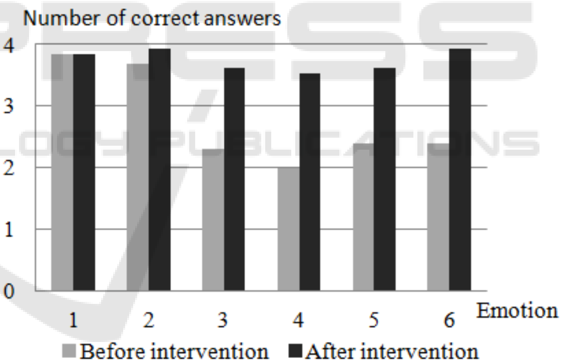

Figure 5: Number of correct answers before and after PC-based intervention averaged over subjects (*n=13*) in the task of emotion recognition in 24 pictures (4 pictures per emotion): 1) happiness, 2) sadness, 3) surprise, 4) fear, 5) anger, 6) disgust.

We did not observe significant difference between two groups of children (PC-based and conventional interventions). The number of correct answers for children with conventional intervention before and after study was 14.88 (7.57) and 21.5 (1.82) respectively.

We also observed that motivation of both groups was approximately the same. Though, the advantage of the tablet PC is in its potential ability to replace trained specialist with less trained one as for example members of children's families. It will definitely make the training technique more widely available and help children to develop highly important skills.

Six months after the intervention subjects' parents were questioned about children's ability to transfer learned skills in their everyday life. Further analysis of the answers revealed that there were 4 groups of children. In the *first group (1 child)* even though after intervention the subject demonstrated progress in 6 months the skills were lost even for the same images as were used during training. In the *second group (4 children)* there was no transfer admitted at all, though they could recognize emotions in the training photos. In the *third group (4 children)* the subjects were able to recognize emotion in their everyday life. For example in cartoon movies or in parents but they used that ability only by request from parents. In the *forth group (10 children)* the subjects used their ability to recognize emotions in everyday life and sometimes they could change their behavior based on the recognized emotion. Some of them were also more easily involved in games than before intervention. Some of them used their ability to recognize emotion only with members of their family but at least 2 of them have changed the behavior with other - not autistic children.

4 CONCLUSIONS

An algorithm for facial images analysis was developed and integrated into emotion perception and production mobile training tools for guiding children in facial expression learning.

Initial clinical study with 19 children with ASD was conducted. After intervention involving the developed tools children's skills were improved and in some cases transferred into their everyday life.

This is the initial milestone toward development of the reliable application that can guide children in their training of perception and production of facial expression without requirements of high level specialist participation, which can make the training tool more widely available.

The future work will focus on development of reliable algorithm for detailed analysis of AU in a wide range of environmental conditions and incorporation of information on head pose into the expression analysis algorithm.

ACKNOWLEDGEMENTS

Work supported by RFH grant № 15-36-01343, SFedU grant № 213.01-2014/001 VG and RFBR grant № 16-31-00384.

REFERENCES

Adolphs, R., Sears, L., & Piven, J., 2001. Abnormal processing of social information from faces in autism. *Journal of Cognitive Neuroscience*, *13*(2), 232–240.

Cockburn, J., Bartlett, M., Tanaka, J., Movellan, J., Pierce, M., & Schultz, R., 2008. SmileMaze : A Tutoring System in Real-Time Facial Expression Perception and Production in Children with Autism Spectrum Disorder. *Computer*, 978–986.

DeQuinzio, J. A., Townsend, D. B., Sturmey, P., & Poulson, C. L., 2007. Generalized imitation of facial models by children with autism. *Journal of Applied Behavior Analysis*, *40*(4), 755–9.

Ekman, P., & Rosenberg, E. L., 2012. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*.

Gordon, I., Pierce, M. D., Bartlett, M. S., & Tanaka, J. W., 2014. Training facial expression production in children on the autism spectrum. *Journal of Autism and Developmental Disorders*, *44*(10), 2486–2498.

Grynszpan, O., Weiss, P. L. (Tamar), Perez-Diaz, F., & Gal, E., 2014. Innovative technology-based interventions for autism spectrum disorders: A meta-analysis. *Autism*, *18*(4), 346–361.

Hashemi, J., Tepper, M., Vallin Spina, T., Esler, A., Morellas, V., Papanikolopoulos, N., … Sapiro, G., 2014. Computer vision tools for low-cost and noninvasive measurement of autism-related behaviors in infants. *Autism Research and Treatment*.

Heimann, M., Nelson, K. E., Tjus, T., & Gillberg, C., 1995. Increasing reading and communication skills in children with autism through an interactive multimedia computer program. *Journal of Autism and Developmental Disorders*, *25*(5), 459–480.

Jain, S., Tamersoy, B., Zhang, Y., Aggarwal, J. K., & Orvalho, V., 2012. An interactive game for teaching facial expressions to children with autism spectrum disorders. In *5th International Symposium on Communications Control and Signal Processing, ISCCSP 2012*.

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M., 2011. The Computer Expression Recognition Toolbox (CERT). *IEEE International Conference on Automatic Face & Gesture Recognition*.

M. Cox, J. Nuevo, J. S. and S. L., 2013. CSIRO Face Analysis SDK. In *AFGR*.

Moore, M., & Calvert, S., 2000. Brief report: Vocabulary acquisition for children with autism. Teacher or computer instruction. *Journal of Autism and Developmental Disorders*, *30*(4), 359–362.

Olszanowski, M., Pochwatko, G., Kuklinski, K., Scibor-Rylski, M., Lewinski, P., & Ohme, R. K., 2015. Warsaw set of emotional facial expression pictures: A validation study of facial display photographs. *Frontiers in Psychology*, *6*(JAN), 1–8.

Wallace, F., & Ekman, P., 1983. EMFACS-7: Emotional facial action coding system. *Unpublished Manuscript, University of California at San Francisco*, *2*(36).