

# Familiar Video Stories as A Means for Children with Autism: An Analytics Approach

Chung-Hao Chen  
Electrical and Computer  
Engineering  
Old Dominion University  
Norfolk, VA USA  
Email: cxchen@odu.edu

Jonna Bobzien  
Communication Disorders  
and Special Education  
Old Dominion University  
Norfolk, VA USA  
Email: jbobzien@odu.edu

Michail Giannakos  
Computer and Information  
Science  
Norwegian University of  
Science and Technology  
Trondheim, Norway  
Email: michailg@idi.ntnu.no

Ann Bruhn  
Dental Hygiene  
Old Dominion University  
Norfolk, VA USA  
Email: abruhn@odu.edu

Alexis Brueggeman  
Computer Science  
Old Dominion University  
Norfolk, VA USA  
Email: abrue005@odu.edu

Shahram Mohrehkesh  
Computer Science  
Old Dominion University  
Norfolk, VA USA  
Email: smohr003@odu.edu

Min Zhang  
Electrical and Computer  
Engineering  
Old Dominion University  
Norfolk, VA USA  
Email: mzhan006@odu.edu

Wei-Wen Hsu  
Electrical and Computer  
Engineering  
Old Dominion University  
Norfolk, VA USA  
Email:  
weiwenhsu@gmail.com

Nikos P. Chrisochoides  
Computer Science  
Old Dominion University  
Norfolk, VA USA  
Email: npchris@gmail.com

**Abstract**—Children with Autism Spectrum Disorder (ASD) may experience extreme anxiety and stress when faced with unfamiliar or novel situations. The use of video social stories is one common approach to helping children with ASD cope with these situations. Despite some literature on the positive outcomes associated with the use of video social stories, the overall results on the success of this intervention are inconsistent. One possible explanation for these inconsistencies is that the characters portrayed in the video are often generic and therefore unfamiliar. In order to address this hypothesis, Chen et al. developed a video face replacement technology, which presents a unique and individualized video to the child that includes familiar faces and places. With this technology, it was found that the personalized videos encouraged children with ASD to express more positive emotions during a novel experience. Although the children with ASD demonstrated decreased stress and anxiety after viewing the video with the face replacement technology, it is still difficult to ascertain how much of the desired information was actually gleaned from watching the video social story. In order to address this critical issue, an existing video learning analytic system was used to determine how subjects with ASD perceived video materials during the viewing of an individualized video social story. The preliminary results of this pilot study demonstrated that subjects with ASD were capable of interacting with the video and did so in order to better understand the content of the video social story. These preliminary results indicate the need for further investigation into how the user interface and its enhancements can provide precise information on the perception and cognition of the individuals with ASD.

**Keywords**—Autism Spectrum Disorders; Social Story; Video Analytics.

## I. INTRODUCTION

According to the American Psychiatric Association [1], Autism Spectrum Disorder (ASD) is a neurodevelopmental syndrome that is characterized by impairments in social communication and in the presentation of repetitive behaviors and restricted interests. Likewise, children with ASD often experience hyposensitivity or hypersensitivity to sensory information in the environment [1]. As such, children with ASD often experience extreme anxiety and stress when faced with unfamiliar or novel situations, such as routine medical visits. Because of this increase in anxiety and stress, the routine visits can become challenging for both the child, who is receiving medical treatment, and the medical professional, who is attempting to administer treatment. In order to reduce the anxiety and fear associated with novel situations, video social stories [2] have been used to assist children with ASD in understanding the new environment. Video social stories are short videos that demonstrate the events that are likely to occur while the child is in a given environment, as well as the expected behaviors of the child during such events. Typically, children with ASD watch the video prior to exposure to the novel event or situation, thereby exposing the child to the environment virtually before they actually experience the environment. Despite some literature on the positive outcomes associated with the use of video social stories, the overall results are inconsistent [3, 4]. A potential reason for the inconsistencies of the success of this technique could be because the children with ASD lose interest and fail to pay attention to the whole video. One possible explanation for this lack of interest is that the characters represented in the

video are unfamiliar; therefore, the children may not be interested in watching. To counteract this potential disinterest, we proposed video face replacement technology, in which the character represented in the video social story was the actual child with ASD who was viewing the video. By watching the individualized video social story, the child with ASD may be able to accurately visualize himself/herself in the novel environment. As a result, the child would be able to feel as though they have already experienced the setting, thereby potentially reducing the anxiety and stress associated with unfamiliar situations.

In our preliminary work [5], we successfully created video face replacement technology that embedded a target child’s lifelike facial features into a video social story. Our preliminary experimental results showed that when the children with ASD were shown character-accurate rendered videos, they were able to engage with the video and comprehend the contextual information presented to some degree. While this study demonstrated that video social stories with replaced faces and adjusted skin colors can effectively attract the attention of a child with ASD, it did not determine how much of the desired information from the video was actually received by the child. The long term goal of this project is to determine if the material presented to the children with ASD was appropriate. As a first step to achieve our goal, in this paper we propose an approach to measure the degree of interaction a child with ASD has with the different part(s) of the video social story. Understanding the “why” is outside the scope of this paper, there might be many reasons we hope to answer in follow up studies. *To the best of our knowledge, there is no -existing engineering tool to measure such data for children with ASD.* In order to address this issue, we adapted a concept from an existing video learning analytics system [6] to track the interaction of children with ASD with video social stories. Employing the proposed user interface and tracking in detailed logs the exact operations (eg. pause, move back or forward, etc.) children with ASD choose to manipulate the video, might help us to understand (in follow up extensive studies from us and others) which parts of the video maintained the child’s interest and attention for different reasons (e.g. more difficult) to comprehend. These identified parts of the video are, therefore, more likely to be learned and have the presented behaviors replicated.

In summary, this work enhances our earlier technology on video learning analytics system [6] by customizing it on subjects with ASD i.e., how they perceive and cognize a given video story. In the future we expect to extend it in other activities related to learning of educational materials. The findings from this system have the potential to offer helpful insights into the preparation of content appropriate materials for interventions and education in autism. This paper is organized as follows: Section II shows the relevant research, the problem we are attempting to solve, and illustrates our concepts. Section III introduces the proposed method in details. The experimental results will be discussed in Section IV.

## II. BACKGROUND

Research shows [3] that children with ASD are extremely anxious and fearful when faced with unfamiliar environments and situations. One conventional solution to facilitate the

understanding of a novel event (e.g., dental visit, medical examination, toilet training, etc.) by children with ASD is to generate a video social story. Through the use of video social stories, children with ASD are exposed to the event as well as the corresponding expected behaviors. Characters in the videos can be an animated subject or a live actor/actress. Despite positive feedback from parents/caregivers regarding the use of traditional video social stories, overall child behavior outcomes are variable. One possible explanation for this variability is that children with ASD may be reluctant to pay attention to the actor’s or actress’s face in the video because this person is unfamiliar and therefore uninteresting.—By drawing upon findings from the psychological theory of self-efficacy -“I know I can because I have done it before” by Bandura [7], a new assistive intervention technology, self-modeling video, has been developed. It is expected that the self-modeling videos can improve upon the outcomes of the conventional video story. For example, the Look at Me Now, LLC., [8] uses the profile photo of a child with ASD to generate a video social story.



Fig. 1. Demonstration of properly rendered images and poorly rendered image from the same situation (Subjects have signed consent forms for media release).

As such, the intention of the video is to encourage the child to feel as if he or she has already experienced the novel situation, thereby potentially reducing anxiety, apprehension, and inappropriate behaviors elicited during the actual event. Again, the outcomes of this video social story format are not consistent due to the attentional difficulties of children with ASD. One possible hypothesis for this inattentiveness by the child with ASD is the quality of rendered images. Our hypothesis was that the photo inconsistencies, such as the white border line around the face and skin pigmentation discrepancy, may distract the child’s attention. Therefore, our first work was to validate this hypothesis [5]. In order to prove our hypothesis, we generated two video social stories with the same contextual information. One video was properly rendered by applying the work of Bitouk et al. [9] to measure the pose, resolution, skin color, and lighting of the subject of interest for replacement in the original image. In other words, we imbedded the child’s lifelike facial expression, skin pigmentation, and head position of the generic child featured in the original video. The other video social story was rendered by applying an approach similar to Look At Me Now, LLC [8]. Fig. 1 demonstrates an example from the two types of video [5]. We tested both videos

on children with ASD by having the children view each video social story and measuring their responses during the video. The experimental results of this preliminary investigation demonstrated that the children with ASD focused their attention more often to the properly rendered video social story than the poorly rendered videos. We found that the child with ASD can better understand what is happening in the properly rendered video social story because they are actually viewing himself/herself in the situation.

Our next interest is to determine how much of the desired information in the video social story was actually transferred to the subjects with ASD. In other words, we want to learn what details and contexts in the actual video are appropriate to facilitate the attention of subjects with ASD. We will adapt the work from an existing video learning analytics system (VLAS) [6] for this study. This system will help us analyze the viewers' knowledge acquisition by indicating to which part of the video social story is the child with ASD most engaged and the correct responses of the respective questions. This will help us better understand how subjects with ASD perceive and recognize the visual stimuli, thus, further helping researchers prepare content-appropriate materials.

The VLAS has been previously used in the field of long-distance education to allow researchers and educators to understand and improve the effectiveness of video-based learning tools and practices. Learners are using video lectures for a variety of subjective and objective benefits, and students perceive video technology as a practical learning resource. However, many aspects related with subjects' a) video-navigation, b) knowledge acquisition, and c) attitudes with the video-assisted tools and methods still remain unexplored:

a) **Navigation** - Are users viewing the entire video?; what segments of the video do students select to view, and why?; how many times do users view any given video segment?; and what video applications are more attractive or engaging?

b) **Knowledge Acquisition** - How do users perform with the assistance of video materials?; is there any relation among users' viewing patterns and the level of the acquired knowledge?

c) **Attitudes** - How do users perceive video-based instruction?; is there any significant shift in their attitudes during a video-based instruction?

To address these critical issues, VLAS by Giannakos et al., [6] provides a first step towards understanding viewers' multi-faceted interactions with video materials. It captures and analyzes viewer's interest based on the viewer's interaction with the video. In order to test how much information from the video has been passed to the viewer (knowledge acquisition), the viewer is asked to answer several questions. The answers to those questions can be found only in the video and cannot be answered with previous knowledge.

### III. PROPOSED METHOD

There are two main steps in our proposed method. The first step is to generate videos from the same video template. One of the videos is generated by our proposed face replacement technology to represent a high quality self-modeling video

social story. The second step is to apply the VLAS [6] to analyze the responses from the individuals with ASD.

#### A. Face Replacement

In order to render a high quality video, we applied the approach based on Bitouk et al. [9] to measure the head position, resolution, color and lighting of the subject, which are used to provide guidance to seamlessly blend the target of interest into the image template. Afterwards, we applied three Gestalt properties (i.e., symmetry, continuity and proximity) to provide feedbacks on the blending quality, which is summarized as following:

Let  $\Omega$  be the domain of definition of a scene image,  $\Omega = R_B \cup R_S$ .  $R_B$  and  $R_S$  denote the regions that belong to backgrounds and foregrounds, respectively. After the background identification, we know that most of the structured objects in the image are contained in a sub-region  $R_S \subset \Omega$ . Let  $P_0$  be the initial partition of  $\Omega$  from a bottom-up segmentation method. Let  $a$  denote a uniform patch from the initial partition  $P_0$ . For  $\forall (a \in P_0) \wedge (a \in R_S)$ ,  $a$  is one of the constituent parts of an unknown structured object. Based on the initial part  $a$ , we want to find the maximum region  $R_a \subset R_S$  so that, the initial part  $a \in R_a$  and for any uniform patch  $i$ , where  $(i \in P_0) \wedge (i \in R_a)$ ,  $i$  should have some special structural relations that obey the non-accidentalness principle with the remaining patches in  $R_a$ . This is formulated as follows:

$$R_a = \arg \min_R (E[\partial R]) \quad \text{with } (a \in R) \wedge (R \in R_S) \quad (1)$$

Where  $R$  is a region in  $R_S$ ,  $\partial R$  is the boundary of  $R$ ,  $E[\partial R]$  is a boundary energy function, which is defined as follows [15]:

$$E[\partial R] = \frac{-\iint_R f(x,y) dx dy}{L(\partial R)} \quad (2)$$

Where  $L(\partial R)$  is the length of the boundary of  $R$ .  $f(x,y)$  is the weight function in region  $R$ . The boundary energy function provides a tool for measuring how good a region is. The criterion of region goodness depends on how the weight function  $f(x,y)$  is defined. One can use  $f(x,y)$  to encode any information (e.g., color or texture) over the region  $R$ .

#### B. Video Learning Analytics System (VLAS)

Studies [10, 11] suggest that students are using video lectures when offered, for a variety of subjective and objective benefits, and those students perceive the video technology as a practical learning resource. As a result, the use of videos for learning has become widely employed in recent years and many educational institutions and business organizations have incorporated video into their instructional materials. However, despite the growing number and variety of video lectures available, there is limited understanding of their effectiveness in terms of how students use and ultimately how they learn from video materials. Thus, the primary purpose of the VLAS proposed by Giannakos et al. [6] is to address this need. Specifically, one of the important questions that the VLAS wants to answer is what segments of the video do students select to view and why, and how much information in the video lecture has been passed to viewers to help them understand the subject. Thus, the study paradigm of VLAS can be extended to

help shed the light on how much desired information in the video has been passed to subjects with ASD. This knowledge may help in the development of more effective video-based interventions and instructions.

Fig. 2 shows how the system was used in our case as well as the user interface for this study. In particular, our participants have the ability to manipulate the video by using a play/pause button and clicking and dragging along the seekbar to move forward and back in the video. Users’ interactions are recorded and stored according to their user unique information. Therefore, an appropriate data analysis was used to identify the interactions of each individual subject. Additionally, to ensure success in operating the system, the subjects were provided with a set of instructions to follow during their viewing time. The instructions included items such as emailing the authors at the beginning and end of the experiment and ensuring to press the start button in order to trigger the system to begin collecting data.

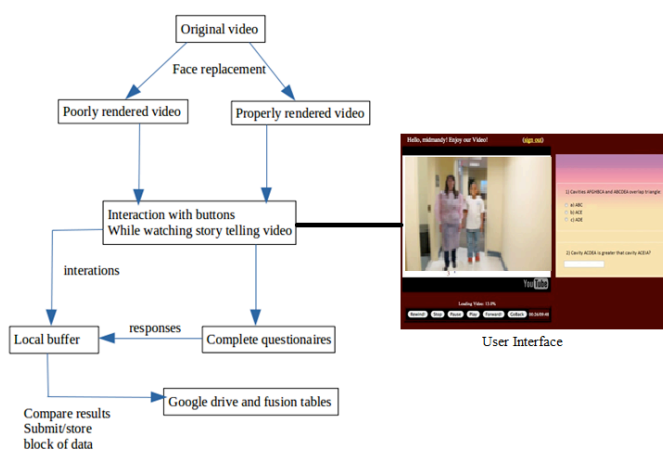


Fig. 2. The architecture of the system and the user interface (Subjects have signed consent forms for media release).

In order to test if the user can properly understand the context in the video and how much information in the video has been passed to him/her, we created two questionnaires based on the story in the video, one requiring open-ended responses and one utilizing a multiple choice format. The questionnaires were available to the subjects throughout their viewing experience. Their answers to the questions were stored and used to analyze whether the context in the video is appropriate. Additionally, in order to identify the attitudes of the subjects with ASD while viewing the video, we used an open-ended survey delivered immediately after the viewing of the video social story. This survey was used to ascertain the opinions of the efficacy of the video social story itself. By doing this, we will learn the relationship between video navigation and the level of cognition, as well as the beliefs of the participants regarding the potential benefits of the video social story intervention. Our preliminary findings are reported in the perception part of the Section V.

#### IV. EXPERIMENTAL PROCEDURES

An initial empirical validation of the proposed idea was implemented. The goal of this empirical validation is to

provide the first analytics-based evidence regarding the effectiveness and acceptability of the proposed concept. The early results should not be seen as a rigorous evaluation of the concept, but as reflections rising from a particular case study as well as empirical evidence for the further development of the idea.

Hence, in our pilot study, two males with ASD were recruited to test the video social story. The two males were both classified as being on the autism spectrum. Subject 1, a 14-year-old teen was considered to be higher functioning than Subject 2, a 12-year-old preteen, because of his advanced social and communication skills. We employed the following data collection methods (measures) for each of the subjects:

**Interactions:** One of the main data collection methods was the viewing interactions by the subjects with ASD; in other words, the interactions of the subjects with ASD with the system (i.e., Play, Pause, Stop, etc.) – video navigation. With the assistance of those interactions among the subjects with ASD and the system (log files) we address questions like, what content did the subjects with ASD watch several times and what content did the subjects with ASD skip?

To be able to investigate this we used VLAS, in VLAS a researcher simply creates a new experiment by adding the URL of the intended YouTube video. Next, the system produces a URL for the experiment which the researcher shares with the study subjects. A screenshot of the URL link when opened in the browser by subjects is illustrated in Fig. 3. Hence, VLAS allows researchers to include questions related to the video and then collect all the analytics related to their video navigation as well as knowledge acquisition. More information about that can be found at the Chorlianopoulos et al. [12].

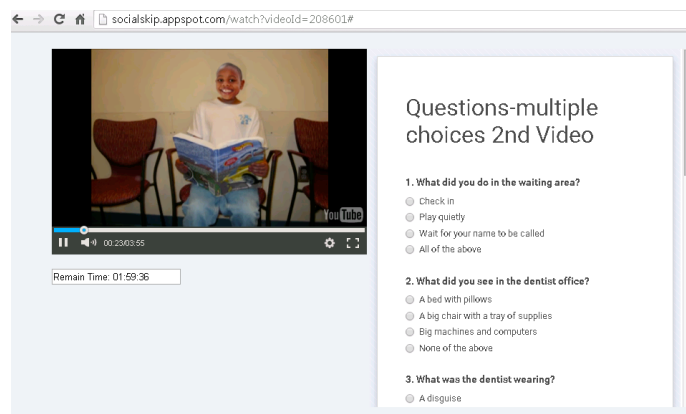


Fig. 3. The interface of the system has familiar buttons, seek bar, as well as questionnaire functionality (The subject has signed a consent form for media release).

**Perception:** Once the viewing of the video has concluded and the subjects with ASD have submitted their responses to the content specific questionnaire, they will report their perceptions and feelings of the process by answering related questions. These questions were generic in nature (i.e., How do you feel when you are using face replacement video?, How would other children feel about watching these videos before

going to the dentist?).

On the next step of this ongoing work we aim to code children responses (e.g., using a qualitative analysis tool like MAXQDA (www.maxqda.com) or NVivo (www.nvivo10.com)); this will allow us to qualitatively understand the potential improvement of our concept.

### V. EXPERIMENTAL RESULTS

**Interactions:** We first evaluated how many interactions each subject had with the VLAS. Subject 1 did not have any interactions (e.g., pause, play, backward, forward) during either episode of watching the videos and answering the questions. In other words, he simply answered the questions as the video played and did not pause, rewind or forward the video. On the other hand, Subject 2 had significant interactions while watching the video and answering the open-ended questions. As viewed in 4, he paused and played the video several times while watching the material. In reviewing video tapes of the sessions, it appeared that Subject 2 used these pause times to read the questions and provide answers. Additionally, Subject 2 forwarded the video four times and rewound the video two times during the initial viewing. However, during the second viewing of the video (i.e. multiple choice questions), Subject 2 did not interact significantly with the video. In fact, he only interacted with the video one time during the first 20 seconds and he did not have any other interactions. Because both Subjects did not seem to need to observe the video to answer the questions during the second viewing, it is assumed that although the videos were viewed a week apart, both Subjects remembered the video content enough to provide responses to the questions. In our future studies we will address this issue by creating more varied contents.

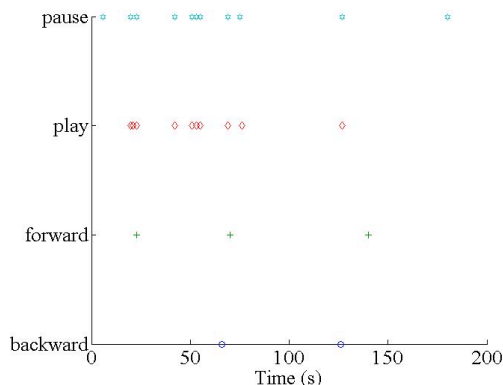


Fig. 4. Time series of interactions by subject 1 at open ended questions experiment

In terms of the number of correct answers to the questions, Subject 1 showed improvement during the second experiment. As shown in Table 1, Subject 1 appeared to use his memory to answer questions. Moreover, since the answers were available as multiple choices, his performance improved (i.e., from 3 to 5 correct answers). However, the performance by Subject 2 actually decreased during the second experience. In this instance, he did not interact with the video which may have contributed to the decreased accuracy of his results.

Table 1. The number of correct answers to 7 questions per subject

	Subject 1	Subject 2
Open Ended Questions	3	7
Multiple Choice Questions	5	5

**Perceptions:** The perceptions of the subjects with ASD regarding the viewing of the video face replacement social story were variable. Both subjects indicated that the video was calming, but that nothing in the video itself stood out as positive or negative. One theme that arose across both participants is the agreement that the video social story would be helpful to children who watched it before going to the dentist (i.e. “It will probably help them to calm down”, “It would be good”). Overall, the perceptions of the subjects with ASD indicated that viewing the video social story would help reduce the anxiety and stress associated with going to the dentist.

### VI. CONCLUSION

Currently, ASD is the most commonly diagnosed neurodevelopmental disability in young children. As such, it is imperative to determine effective strategies and interventions that may help to reduce the stress and anxiety associated with novel events. Through the use of an existing video interface, this pilot study demonstrated that participants with ASD were able to successfully navigate a video social story as demonstrated by positive outcomes on measures of the video content. Furthermore, the participants with ASD were effectively able to perceive and cognitively understand the video content. These results indicate the need for further investigation into the use of video interface to determine the effectiveness and efficacy of video face replacement social stories.

### VII. ACKNOWLEDGMENTS

This study in part is funded by NSF grant CCF-1439079 and the Richard T. Cheng Endowment. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NSF. We want to thank for all the people worked for the video analytics system (socialskip.org)

### REFERENCE

- [1] American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5), American psychiatric Publishing, 5th Edition, 2013.
- [2] C. Gray, “Social stories 10.0: the new defining criteria”. *Jenison Autism Journal*, 15, 1–21, 2004.
- [3] E. O’Conner, “The use of social story DVDs to reduce anxiety levels: A case study of a child with autism and learning disabilities. *Support for Learning*, 24(3), 133-136. 2009.
- [4] D. Scattone, M. Wilzenski, R.P. Edwards, and N.B. Rabia, “Decreasing disruptive behaviours of children with autism using social stories.” *Journal of Autism and Developmental Disorder*, 32, 535–543. 2002.
- [5] C. Chen, J. Bobzien and A. Bruhn, “The first step toward reconstructing visual perception and cognition in autism”, 2014 IEEE International Conference on SMC, pp. 1171 – 1174.
- [6] M. Giannakos1, K. Choriantopoulos, and N. Chrisochoides, “Making sense of video analytics: lessons learned from clickstream interactions,

attitudes, and learning outcome in a video-assisted course”, *International Review of Research in Open and Distance Learning* 02/2015; 16(1):260-283.

- [7] A. Bandura, *Self-efficacy: the exercise of control*, New York: Freeman, 1997.
- [8] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar, “Face swapping: automatically replacing faces in photographs”, *ACM Transactions on Graphics (SIGGRAPH)*, August 2008.
- [9] Look at Me Now!, LLC. Retrieved April 22, 2015, from <http://lookatmenow.org>.
- [10] A. Chester, A. Buntine, K. Hammond and L. Atkinson, “Podcasting in education: student attitudes, behaviour and self-efficacy”, *Educational Technology & Society*, 14(2), 236–247, 2011.
- [11] R. Van Zanten, S. Somogyi, and G. Curro. “Purpose and preference in educational podcasting”, *British Journal of Educational Technology*, 43, 1, 130–138. 2012. doi:10.1111/j.1467-8535.2010.01153.x.
- [12] K. Chorianopoulos, M. N. Giannakos, N. Chrisochoides. “Open system for video learning analytics”, *Proceedings of the First ACM Conference on Learning*. pp. 153-154, 2014.