EMOTIONS AND EMOTICONS:

FACIAL EXPRESSION RECOGNITION APP ACCURACY

by

Anna Stewart

HONORS THESIS

Submitted to Texas State University
in partial fulfillment
of the requirements for
graduation in the Honors College
May 2020

Thesis Supervisor:

Maria D. Resendiz

Second Reader:

Amy Louise Schwarz

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

This work is protected by the Copyright Laws of the United States (Public Law 94-553, section 107). Consistent with fair use as defined in the Copyright Laws, brief quotations from this material are allowed with proper acknowledgement. Use of this material for financial gain without the author's express written permission is not allowed.

Duplication Permission

As the copyright holder of this work I, Anna Stewart, authorize duplication of this work, in whole or in part, for educational or scholarly purposes only.

ACKNOWLEDGEMENTS

I am immensely grateful and thank God for all that have supported me as I completed this project. I would especially like to thank my supervisor, Dr. Maria Resendiz. Over the last three years, she has invested her time and wisdom into me. I would like to thank my second reader, Dr. Amy Louise Schwarz, for her encouragement and teaching me how to navigate the research process with intellectual good faith. This study is a part of a larger interprofessional education project and would not be possible without the mASkeD app creators: Dr. Damian Valles, Inzamam Haque, and Rezwan Matin in the High-Performance Engineering Research Group. Lastly, I want to thank my parents and sister, whose love and support gave me strength during this process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

LIST OF ABBREVIATIONS

Abbreviation	Term
ASD	Autism Spectrum Disorder
AAC	Augmentative Alternative Communication
AI	Artificial Intelligence
ANN	Artificial Neural Network
DNN	Deep Neural Network
COVID – 19	Coronavirus
CNN	Convoluted Neural Network
SLP(s)	Speech-Language Pathologist(s)
FER-2013	Facial Expression Recognition Dataset 2013
KDEF	Karolinska Directed Emotional Faces Dataset

ABSTRACT

Although technology exists that can identify emotions, existing technology is expensive, not a stand-alone product, and was not created as an intervention tool for individuals with Autism Spectrum Disorder (ASD). There is a gap in available apps that can generalize facial feature information and accommodate for lighting and angle variance. A graduate student at Texas State University created an app for real-world use that addresses this gap by utilizing deep machine learning to identify emotions in various lighting conditions and from different angles. This iOS application detects a human face, processes the image with deep machine learning, identifies the facial expression, and shows a corresponding emoticon representing the emotion identified.

This pre-pilot study aims to identify the level of agreement between people and the emoticon generated by the mASkeD app. Ten participants completed a 44 question survey including screenshots of volunteers demonstrating emotions in real-time and emoticons generated by the mASkeD app. Participants were asked the question “Do you agree with the emoticon generated by the mASkeD app?” and responded with one of the following responses: yes, no, or I don’t know.

The overall percentage of agreement between people and the emoticon generated by the mASkeD app was relatively low. There were some emotions that had higher levels of agreement between people and the emoticon generated by the mASkeD app. Since this is the first time the mASkeD app has been piloted in real-time, the results provide support for continued development and refinement.

I. LITERATURE REVIEW

The long-term objective of the larger study is to develop an app for individuals with Autism Spectrum Disorder (ASD) for use as an intervention tool as part of speech-language therapy (Resendiz & Valles, in preparation). The first step to creating an app that can mimic human emotion identification for those with ASD is to assess agreement for the visual portion of the mASkeD app among neurotypically developing individuals.

Universal Facial Expressions of Emotion

Non-verbal communication modalities (facial expression and body gestures) are powerful agents in social interaction. Infants who are following a typical course of development discriminate between happy, sad, and surprise static faces at 3-4 months old (Young-Browne et al., 1977). Non-verbal cues enable people to communicate on a fundamental level despite their language background. This is important for the expression and interpretation of ideas, thoughts, and feelings (Burgoon et al., 2016).

Specifically, facial expressions play a significant role in social exchanges and give us insights into the emotional states of those around us (Frith, 2009). Happiness, sadness, anger, fear, surprise, and disgust (Ekman & Friesen, 1971) are facial expressions of emotions that are identified as pan-cultural or universal. These expressions are identifiable by people from cultures and communities that differ in education level, language, and age (Ekman & Friesen,1971). While for most, the expression and identification of emotions seems to be an easy skill to acquire, there are populations that have great difficulty with facial expression recognition.

Facial Expression Recognition & Individuals with ASD

Initially characterized as a disturbance of affective contact (Kanner, 1943), pragmatic difficulties often mark ASD. Diagnostic criteria for ASD include deficits in non-verbal communication, emotional-social reciprocity, and behavior adaption for various social contexts (American Psychiatric Association, 2013). A meta-analysis covering a wide range of ages, emotion recognition tasks, and IQ scores showed that individuals with ASD have difficulties with emotion recognition (Uljarevic & Hamilton, 2013).

Speech-Language Pathologists (SLPs) are therapists that specialize in communication and helping individuals with ASD communicate to be successful in their daily social activities (ASHA, 2020). Specifically, SLPs work collaboratively with other health care professionals to diagnose ASD and provide pragmatic treatment (ASHA, 2020). SLPs teach people with ASD how to communicate, sometimes using Augmentative and Alternative Communication (AAC). SLPs establish treatment dismissal criteria by determining when people with ASD can effectively communicate in daily social situations independently (ASHA, 2020).

Artificial Intelligence

Within the last few decades, the technology industry has posed the question: Can computers be trained to recognize facial expressions of emotions? Artificial intelligence (AI) is the study of the technological developments that can complete "human-like processes like learning, reasoning, and self-correction” (Kok et al., 2009). Engineers and data scientists can utilize AI to create facial expression recognition technology (Figure 1).

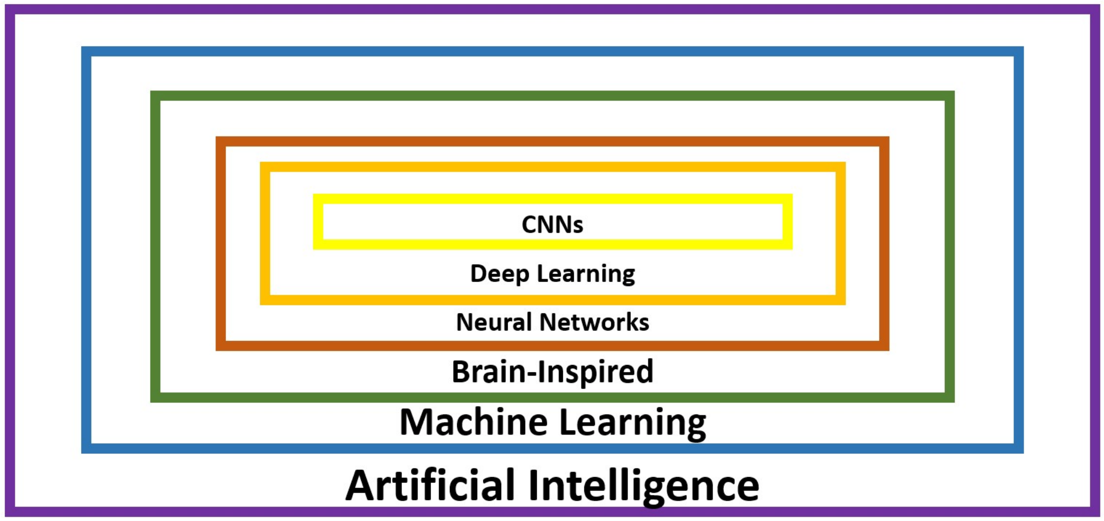


Figure 1. Pathway to Computers Identifying Emotions

One *approach* to achieving AI for emotion recognition utilizes machine learning (Agrawal, 2018). Machine learning focuses on making computers readjust and revise their actions for greater function accuracy (Marsland, 2015). This enables computers to learn without explicit programming (Samuel, 1959). However, the traditional machine learning model could not directly input information (like an image) into an algorithm- it requires feature extraction algorithms (Haque, 2019). Feature extraction algorithms require an input image to be quantified and translated into a list of numbers (vectors) that quantifies the contents of the image. All these steps must occur before the information is put into a machine learning algorithm (Haque, 2019). There is a sub-area of machine learning that is inspired by human brains called Artificial Neural Networks (ANNs) (Argawal, 2018). Similarly, to how neural pathways connect varying information in the brain, ANNs form an algorithm of data connections called layers. ANNs consist of input layers, output layers, and hidden layers. Hidden layers are connections created in between input and output layers. When these networks include more than one hidden layer, the network is considered deep (Figure 2) (Haque, 2019). Deep neural networks are used in deep learning.

Deep learning is a specialized and sophisticated *technique* of machine learning (Agrawal, 2018). For example, for object recognition, layers can represent an area of the human face. One hidden layer may identify the edges of a face, while another layer may identify the contours on a face. Hidden layers are hierarchically arranged to assess a complete image (Haque, 2019). Each layer is connected to the previous layer and automatically learns the collective features. Deep Neural Networks (DNN) can have higher performance than traditional machine learning. Higher performance can refer to higher accuracy and efficiency.

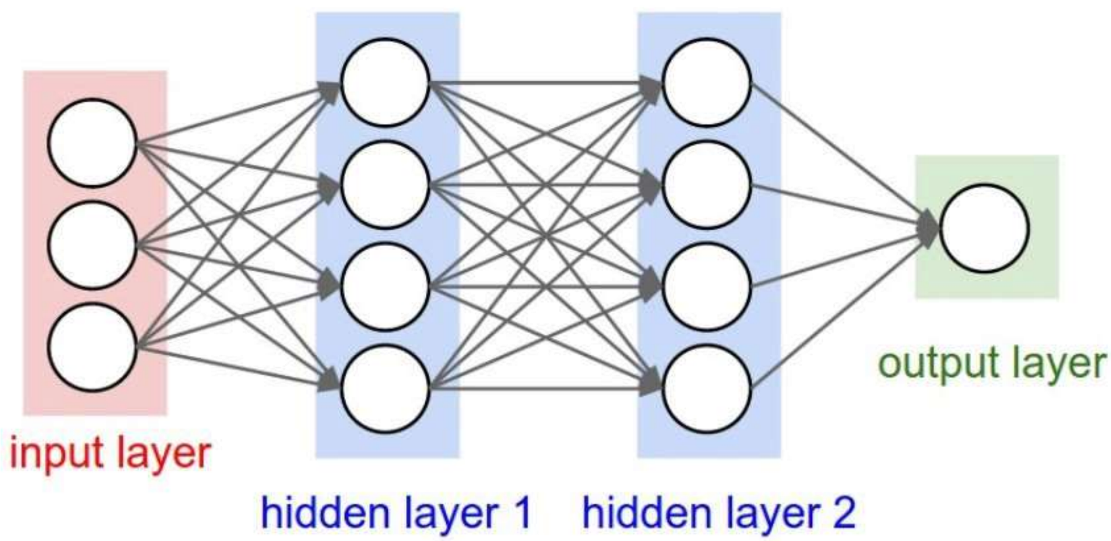


Figure 2. Graphical Representation of a DNN (Stanford, 2020)

For a computer to identify the facial features in an image, it must have computer vision, or the ability to identify objects in an image. The best option for classifying images using computer vision involves a specialized deep learning algorithm called convoluted neural networks (CNNs) (Haque, 2019). CNNs can filter and categorize facial features in images. For more detailed information on CNNs (see Haque (2019).

Existing Technology

Currently, three platforms have designed technology that can identify human emotions using deep learning instead of the traditional machine learning feature-matching: Affectiva, Google Glass and the Autism Glass project, and the Kaggle Challenge. In this section, each technology is described and assessed to determine whether the technology could be used by an SLP as a therapy tool for helping clients with ASD identify emotions in real-time.

Affectiva. Affectiva is a human perception AI software company. The purpose of this company is to use their large emotion repository, computer vision, and deep learning to advance their AI in the following areas: media and audience analytics, automotive AI, and biometric research. Affectiva has the world's largest emotion repository, with over 4 billion frames and 75 countries represented (Zijderveld, 2017). This repository creates an extensive "normative database, a benchmark of what responses to expect in each region of the world" (Zijderveld, 2017). The facial repository trains and tests their facial expression recognition algorithms. Their algorithms utilize multi-modal deep machine learning to analyze both facial and vocal information. Affectiva claims accuracy in the high 90th percentile (Affectiva, 2017).

Affectiva works with ridesharing companies to monitor the emotional and cognitive states of drivers for road safety. Affectiva also works with marketing companies to determine media spend, test brand reveals, and test movie trailers (Affectiva, 2020). Their software, when combined with biometric sensors, helps universities and companies complete human behavior research. Research areas including medicine, ASD, and computer science use Affectiva's software (Affectiva, 2020).

While this company has access to an extensive normative database and uses deep learning to identify emotions, their services are costly. Access to their technology and biometric sensors for research costs a minimum of ten-thousand dollars. Moreover, while other institutions use their technology for research in human behavior, including ASD, they have no technological solutions or interventions specifically created for individuals with ASD.

Google Glass. Initially, created to provide users a hands-free Android system experience, Google Glass is wearable technology that can make everyday tasks easier and quicker. Project Glass focused on making phone and video calls, scheduling appointments, providing interactive GPS directions, and even checking in to restaurants and events (Google, 2012). The latest enterprise edition 2 focuses on professional use in factories, warehouses, and hospitals for organization and workplace efficiency (Google AR & VR, 2019). This hands-free device shows information in the peripheral vision of employees to make hands-on tasks easier and faster.

The Wall Lab at Stanford University developed the Guess What? app that uses AI to analyze the behaviors of children while interacting with their family. This information is used in research about developmental delays in children (Stanford University, 2019, *Guess What Overview*).

Stanford University coopted Google Glass technology and their machine learning software for research on children with ASD in the Autism Glass Project. The purpose of this project was to research socialization and emotion recognition solutions for children with ASD who have limited access to intensive behavioral therapy (Stanford University, 2019, *About the Project*). Autism Glass uses the front-facing camera of Google glass and machine-learning software to help children with ASD recognize emotions.

While this technology has shown promising evidence for improving socialization skills of children with ASD, this technology is not a standalone product. It requires both Google Glass and a smartphone. Google Glass is also expensive with the latest model's starting cost at $999 (Haselton, 2019). SLPs, specialists in communication disorders who work with children with ASD, were not included as part of the Google Glass intervention development team.

Kaggle challenge. In 2013, Kaggle, an online data scientist community, presented the "Challenges in Representation Learning: Facial Expression Recognition" contest (Kaggle, 2013). The purpose of this contest was to discover the new advancements in "representation learning, with a special emphasis on testing the capabilities of current representation learning algorithms and pushing the field towards new developments" (Goodfellow et al., 2013).

With 56 teams, 63 competitors, and 190 entries participating in the contest, the top test accuracy score was approximately 71.16%. The competitors used an image dataset created by Google images called the Facial Expression Recognition dataset (FER-2013). The FER-2013 dataset contained 48x48- pixel, grayscale, front-view photos. Also, this dataset did not include photos from varying lighting conditions (Haque, 2019). This model's lack of training in varying facial angles and conditions limits its real-world applicability in emotion recognition.

Limitations in Current Technology. Within the current technology discussed above, there is no available stand-alone, affordable technology that identifies facial expressions of emotions, accommodates both lighting and angle variance, and is created as an intervention tool for individuals with ASD as part of speech-language therapy with SLPs.

mASkeD app

Developed by a graduate student at Texas State University, the mASkeD app is an application created for Individuals with ASD as an intervention tool for recognizing facial expressions (Haque, 2019). The mASkeD app is designed to be used as part of speech-language therapy with an SLP. The mASkeD app uses deep CNNs, is trained to accommodate different light and angle conditions, and will be freely available. Once opened, the app activates the rear camera. The user then points the camera in the direction of their communication partner's face. An emoticon, representing the communication partner's facial expression, appears at the bottom of the screen (Figure 3). The mASkeD app was trained to identify the following emotions: happy, sad, fear, surprise, anger, disgust, and neutral. This application was tested with still photos and determined to have 75% accuracy (Haque, 2019).

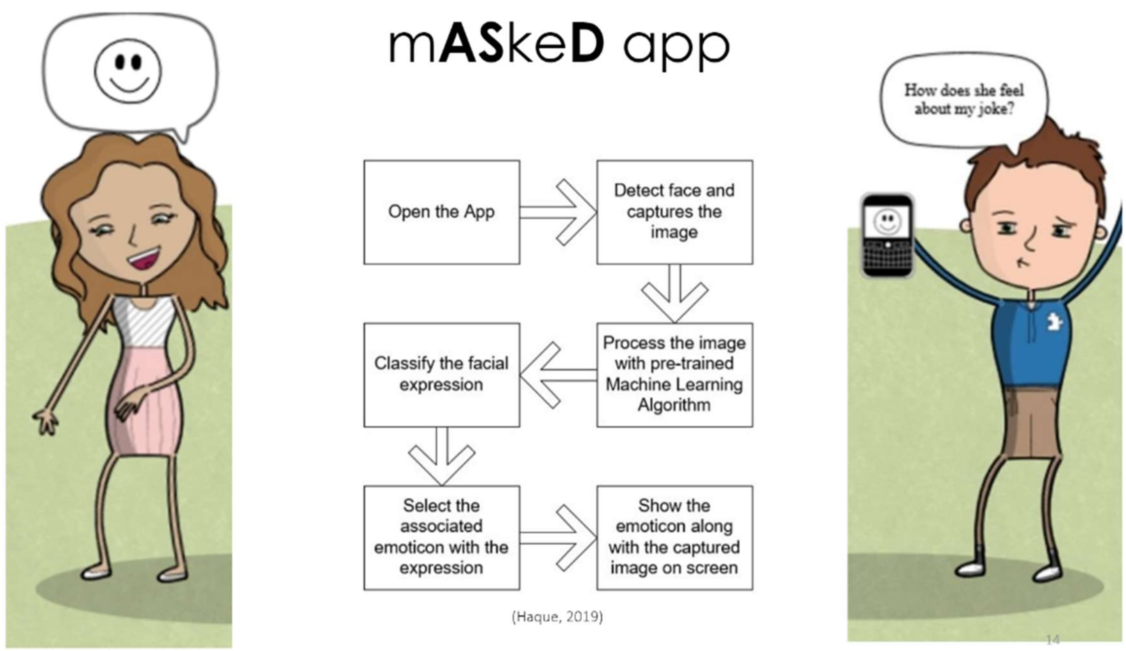


Figure 3. How the mASkeD app Works

Training. The mASkeD app was trained with the FER-2013 and the Karolinska Directed Emotional Faces (KDEF) dataset. Both training datasets included still photos. While the KDEF dataset contained an equal number of images for each emotion (700), the FER-2013 dataset did not. The combined total amount of training images for each emotion included in the training sets were happy (9689), neutral (6898), sad (6777), fear (5821), anger (5653), surprise (4702), and disgust (1247). After the training process, the mASkeD app was tested and determined to have above 75% accuracy in varying lighting conditions and from different angles with still photos (Haque, 2019).

The creation of the KDEF dataset included 70 amateur actors without facial hair, jewelry, makeup, or eyeglasses (Lundqvist et al., 1998). The actors were given one hour to rehearse the facial expressions before taking photos. They were told to make the expressions "strong and clear" (Lundqvist et al., 1998). The conditions in which the photographs for the KDEF dataset were collected do not generalize well to the real life interactions that people with ASD face in their daily efforts to interpret emotions.

L**ong-term goal.** The long-term goal of the larger study is to create an app for use by people with ASD in everyday social situations. The app will utilize video and audio streaming to allow for the identification of emotions using three components for communicating emotions: facial expression, voice intonation, and body language. One phase of development will include focus groups with people with ASD, keeping the end-user in mind. Before the product is presented to focus groups including people with ASD, the development team wants to create a polished product that includes all three components. The mASkeD app will be used in real-world settings so that SLPs can have an intervention tool for individuals with ASD.

For the mASkeD app to perform well in real-world settings, the app must identify emotions of people in both well and poorly-lit environments. The application must identify the emotions of people who are not looking directly into the camera of the device. Before the mASkeD app can be used as an intervention tool for individuals with ASD, each component of the app (facial expression, voice intonation, and body language) must first be evaluated with individuals who are neurotypically developing. The goal of the mASkeD app is to mimic humans in their identification of emotions.

Initial Stage. With the long-term goal of the mASkeD app in mind, the initial stage of this project was to assess the facial expression component with screenshots of the mASkeD app being used in real-time. The screenshots were captured from non-actors (volunteers) demonstrating facial expressions of emotions in varying lighting conditions from different, or natural, angles. This pre-pilot study was completed in the form of a survey to evaluate the agreement of people who are neurotypically developing with the emoticon generated by the mASkeD app.

Summary and Research Question

The long-term objective of a larger study is to develop the mASkeD app for individuals with ASD that can be used as part of speech-language therapy with the support of SLPs (Resendiz & Valles, in preparation). The first step to creating an app that can mimic human emotion identification for those with ASD is to assess agreement for the visual portion of the mASkeD app among neurotypically developing individuals.

The current pre-pilot study aimed to answer the following question: Do people who are neurotypically developing agree with the emoticon generated by the mASkeD app? This will be measured by participants' level of agreement with screenshots of volunteers demonstrating the emotions in real-time and emoticons generated by the mASkeD app.

II. METHOD

The evaluation of agreement in facial expression identification was analyzed using a Qualtrics Survey. Stimuli were developed using screenshots of the application with thirty-six real-time images. Participants who completed the survey saw images such as the picture in Figure 4.

Figure 4. Screenshot of the mASkeD app Displaying an Emoticon

Procedures

Stimuli Development

App installation and Screenshots. The mASkeD app can run on any Apple device with iOS version 11 and beyond (Haque, 2019). The mASkeD app was installed onto one iPad and one iPhone, running an iOS version 11 or higher. Installing the mASkeD app involved installing XCode onto a MacBook laptop, downloading the code contents, and side-loading the app onto the devices. The mASkeD app is not currently available on any app store to download by the public. Six volunteers (two males, four females) ranging in age from 2 to 72 participated in the stimuli development (Table 1).

AGE	GENDER	ETHNICITY	Number of Photos	Emotions
2	Male	Hispanic	2	Fear, Happy
3	Female	Hispanic	4	Fear, Happy, Anger, and Sad
5	Female	Hispanic	4	Happy (2), Sad (2)
39	Female	Hispanic	13	Surprise (2), Fear (3), Happy, Sad (2) Disgust (2), Anger, Neutral (2)
41	Female	Hispanic	5	Fear, Happy, Disgust, Angry, and Neutral
72	Male	Hispanic	6	Fear, Happy (2), Disgust, Neutral, and Sad

Table 1. Volunteer Demographic Information

First, volunteers stood in front of the Apple devices and demonstrated the following expressions: happy, sad, anger, disgust, surprise, fear, and neutral. Volunteers were instructed to demonstrate the emotions, but they were not instructed to look directly at the camera or to modify the pre-existing lighting conditions of their environment. While the volunteers demonstrated the expressions, the app identified the expressions and showed a corresponding emoticon at the bottom of the screen. Then, the volunteers demonstrating emotions in real-time and the emoticon generated by the mASkeD app were recorded through screenshots. There was a total of 36 screenshots. These screenshots served as the stimuli for the survey.

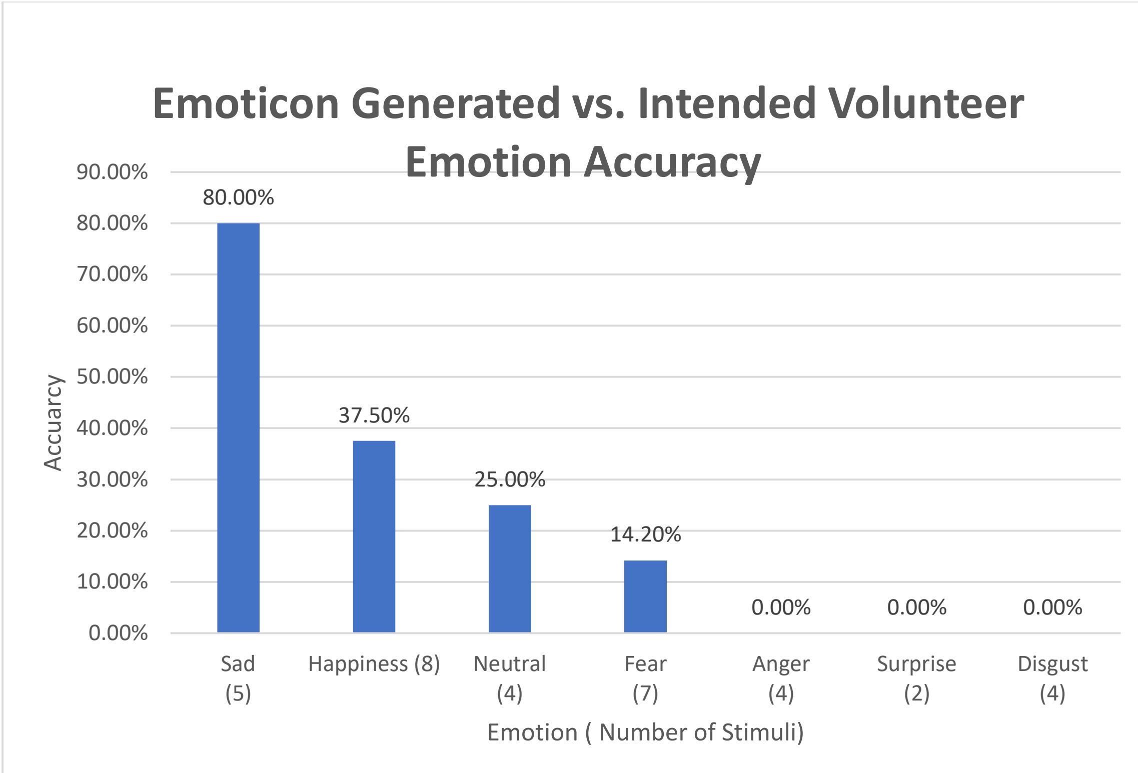


Figure 5. Emoticon Generated vs. Intended Volunteer Emotion Accuracy

Intended Volunteer Emotion vs. App Identification. The research question for the current pre-pilot study did not involve the communication partner's (volunteers') intended emotion. However, it is important to note the agreement between the intended volunteer emotion and the emoticon generated by the mASkeD app was low (Figure 5). The intended volunteer emotion for which agreement was highest was sad, while the intended volunteer emotions for which agreement was lowest, at 0%, were anger, surprise, and disgust.

Participants. All participants were undergraduate students, 18 years of age or older, and had no known psychological diagnosis. The target size for this study was 25 participants. Recruitment information was sent to potential participants via email (See Appendix A).

In total, 89 emails were distributed with ten completed surveys returned, accounting for an 11.2% rate of participation (Table 2). Participants had the option to receive an Amazon gift card after completing the survey. If they chose to receive the gift card, they provided the researchers with an email that was not associated with their responses. These emails were sent a link to claim a $10 e-gift card.

Table 2. Participant Demographic Information

Age	Ethnicity		Gender		Average Number of Languages Spoken (Range)	Education Level
18-36	Hispanic or Latino (9)	Caucasian (1)	Female (9)	Male (1)	1.9 (1-3)	Some College – Bachelor's Degree

Qualtrics Survey

A Qualtrics survey was created to evaluate the agreement between neurotypically developing people and the emoticon generated by the mASkeD app. The survey was distributed via email with an anonymous, reusable link. The survey did not record identifiable information (like the IP addresses) of participants. With a total of 44 questions, the survey consisted of three question types: demographic, self-report, and agreement tasks.

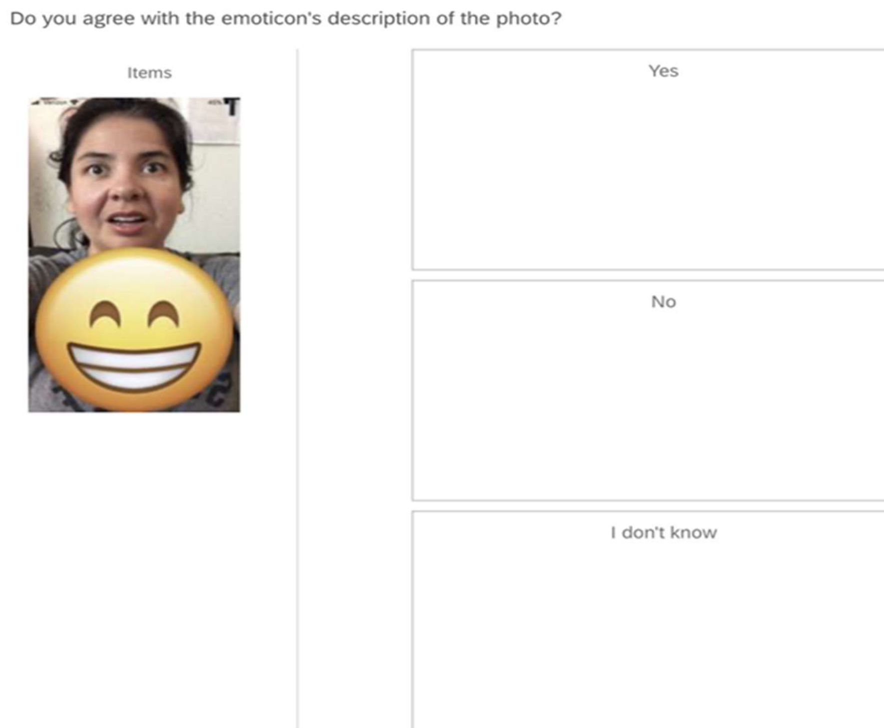


Figure 6. Sample Agreement Task

This study included demographic questions to consider variables that might contribute to findings. Questions were in multiple-choice format and included the following information: age, ethnicity, gender, level of education, and the number of languages spoken. The agreement tasks (Figure 6) contained images of the 36 screenshots collected during stimuli development. Participants were asked, "Do you agree with the emoticon's description of the photo?" and chose one of the following responses: yes, no, or I don’t know by moving the photo on the survey into one of three boxes (yes, no, I don’t know). Participants did not receive feedback on their responses during or after their completion of the survey. I don’t know was included as an answer choice for the agreement tasks to prevent forced responses. If participants had a high rate of uncertainty in the emotion presented in the screenshot photo, this could have implications for the mASkeD app's overall accuracy as well as accuracy regarding emotions that are more difficult to differentiate (e.g., anger vs. disgust, happy vs. surprise) (Ekman & Friesen, 1971).

III. RESULTS

To determine if neurotypically developing people agreed with the emoticon generated by the mASkeD app, participants completed a survey. They chose yes, no, or I don't know when asked if they agreed with an emoticon generated by the mASkeD app.

One screenshot did not upload correctly into the survey, and another screenshot photo looked similar to another photo - these stimuli were removed from the data analysis. The removal of the two screenshot photos followed the procedures of Goodfellow et al. (2013), who removed duplicate images from the FER-2013 dataset. We removed the two stimuli from the data analysis.

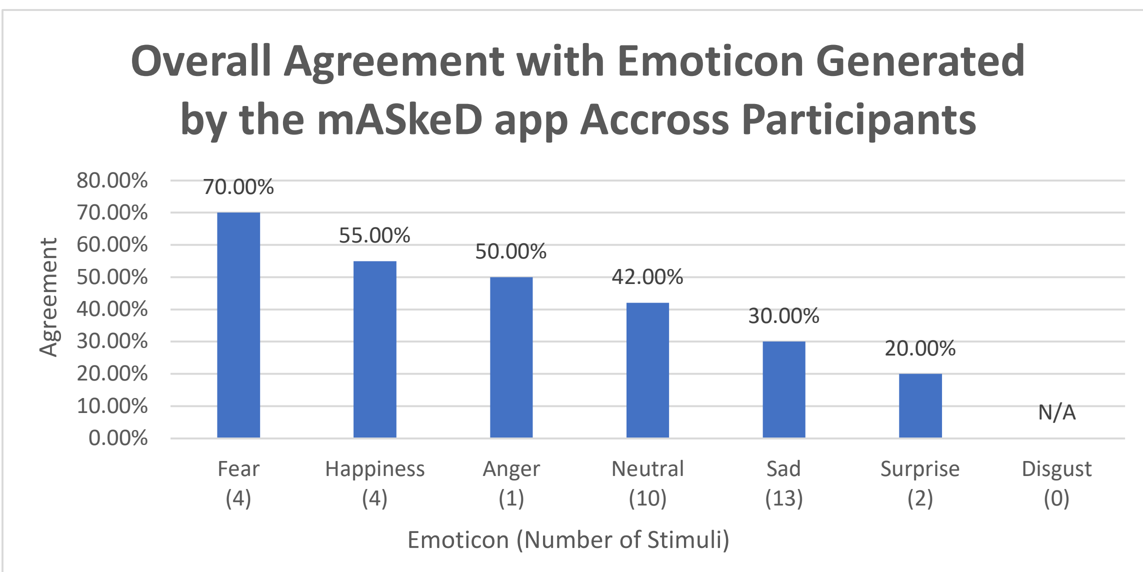


Figure 7. Overall Agreement with Emoticon Generated by the mASkeD App Across Participants

Figure 7 depicts the overall participant agreement with the emoticon generated by the mASkeD app. There were different amounts of agreement for different emotions. The fear emoticon had the highest percentage of agreement with 70%, and the surprise emoticon had the lowest percentage of agreement at 20%. The mASkeD app never generated the disgust emoticon. The emoticons generated were not evenly distributed across the seven emotions.

Because the mASkeD app was previously tested and determined to have 75% accuracy with still photos (Haque, 2019), further analysis was completed on the screenshots to determine which emotions were represented in the 70-100% agreement range (Figure 12). Figures 8-11 show the screenshots that had 70-100% agreement across participants. The caption under each photo represents the target emotion of the volunteer. Captions containing an asterisk show agreement between the emoticon and the volunteers' intended emotions. For example, for the first picture in Figure 8 the intended emotion of the volunteer was happy and the emoticon for happy was generated by the mASkeD app.

There was also 100% agreement across all participants with the happy emoticon generated by the app for the first picture in Figure 8. The last picture in the row also had 100% agreement across participants; however, the volunteer's *intended* emotion was surprise and the emoticon *generated* by the mASkeD app was fear. Remember that the purpose of the mASkeD app is to mimic human identification of emotions, not to read the mind of the communication partner (volunteer); the emotion one is trying to portray and the emotion that others observe can be different. It is important to note that the volunteer photos in this pre-pilot study are not like the photos used in the KDEF dataset. The volunteers in this study were not given an hour to practice facial expressions before stimuli creation. The goal of the larger study is to create an app that can be used in real-time and in real-world environments. The volunteer photos were intended to extend a situation approaching real life interactions with the goal of testing the mASkeD app's generalizability.

Figure 8. 100% Agreement Across Participants

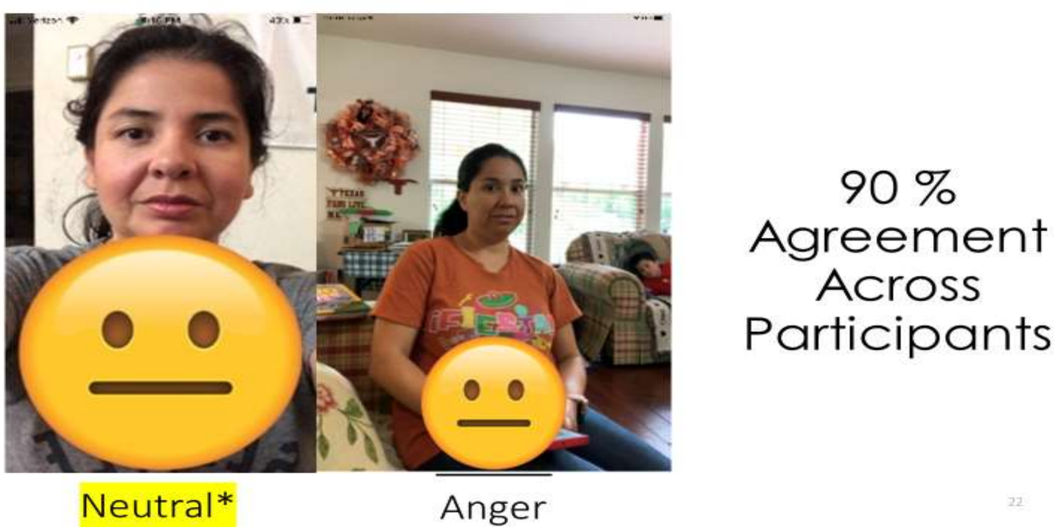


Figure 9. 90% Agreement Across Participants

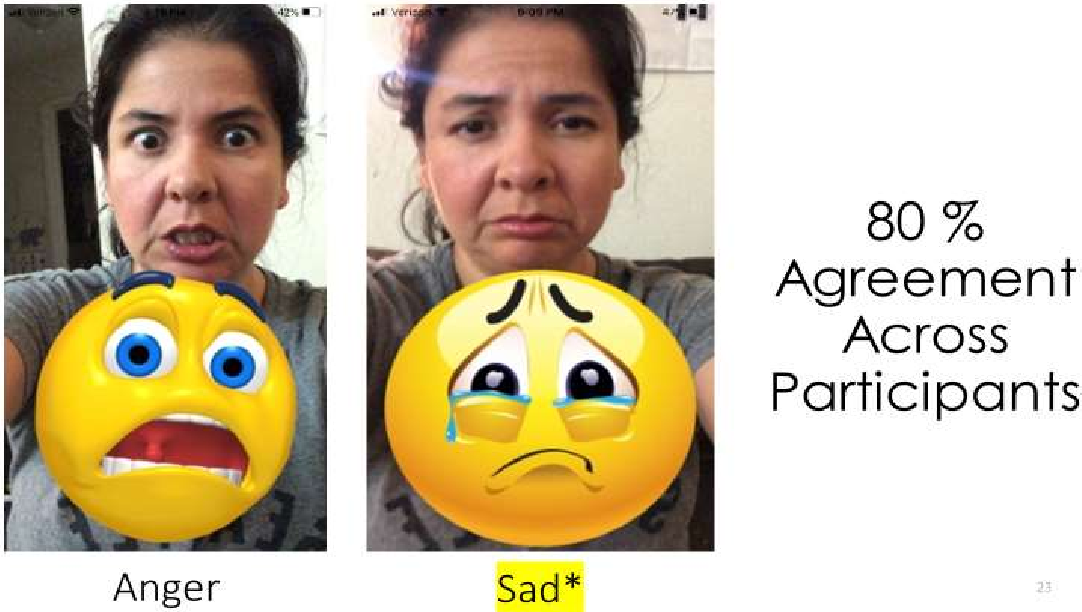


Figure 10. 80% Agreement Across Participants

Figure 11. 70% Agreement Across Participants

Higher Levels of Agreement Between People and Emoticon

Eleven of the 36 (30.5%) stimuli had 70-100% agreement between people and the emoticon generated by the mASkeD app. For the 11 stimuli for which agreement was 70-100%, there were two stimulus photos with the happy emoticon, 3 stimulus photos with the neutral emoticon, 3 stimulus photos with the sadness emoticon, and 3 stimulus photos with the fear emoticon. There were relatively high levels of agreement across these 4 emoticons generated by the mASkeD app when evaluating the stimuli for which there was 70% or greater agreement across participants (Figure 12).

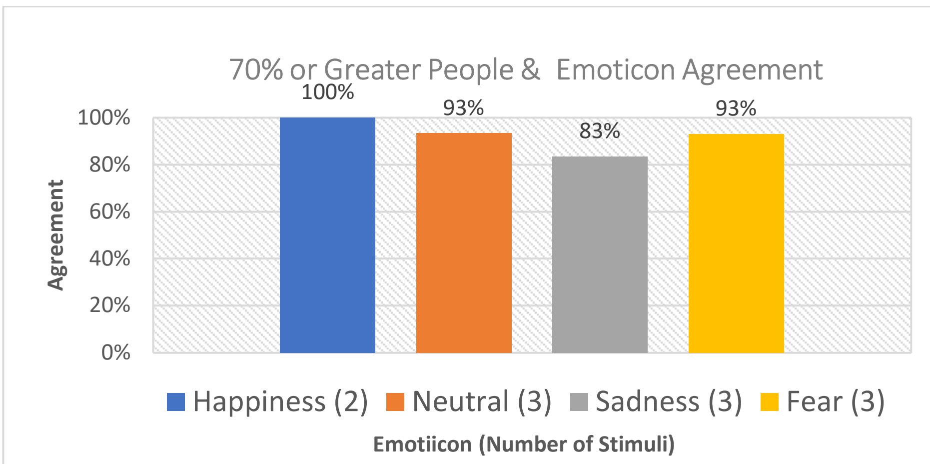


Figure 12. 70% or Greater People & Emoticon Agreement

Agreement Between Intended Volunteer Emotion and Emoticon Generated. Nine of 34 (26.5%) emoticons generated by the mASkeD app agreed with the volunteers' intended emotions. Of these 9 emoticons generated by the mASkeD app, 7 (77.7 %) of the emoticons had 70% or greater agreement between participants and the emoticon generated by the mASkeD app. The emotions included: happy, sad, fear, and neutral (Figure 13).

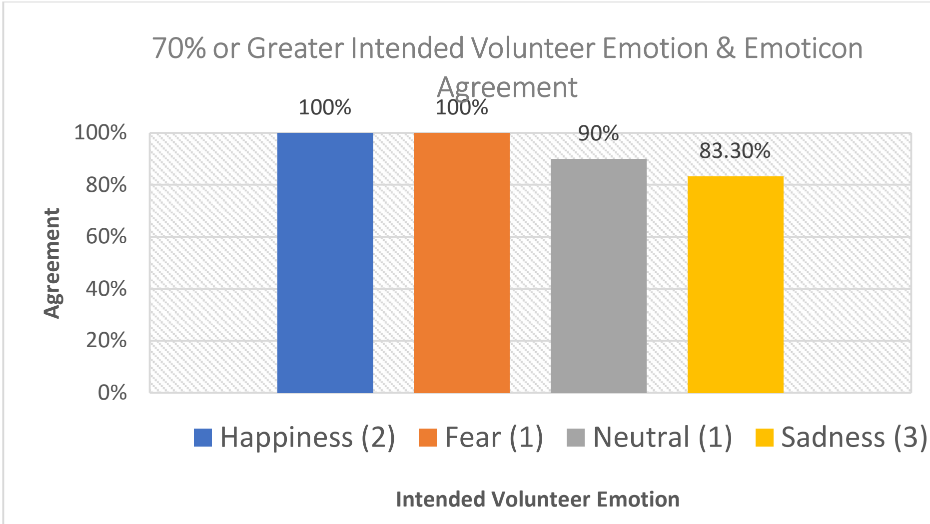


Figure 13. 70 % or Greater Intended Volunteer Emotion & Emoticon Agreement

Intended Volunteer Emotion and Different Emoticon Generated. Twenty-five of 34 (73.5%) emoticons generated by the mASkeD app did not agree with the volunteer's intended emotion. Of these 25 emoticons generated by the mASkeD app that did not agree with the intended volunteer's emotion, 4 (16%) of the emoticons had 70% or greater agreement between participants and the emoticon generated by the mASkeD app. Intended emotions for which there was a lack of agreement between the volunteer's intended emotion and the emoticon generated by the mASkeD app, but for which there was agreement between people and the emoticon generated included: angry, surprise, and fear (Figure 14). Recall that these were only 4 of the stimulus photos.

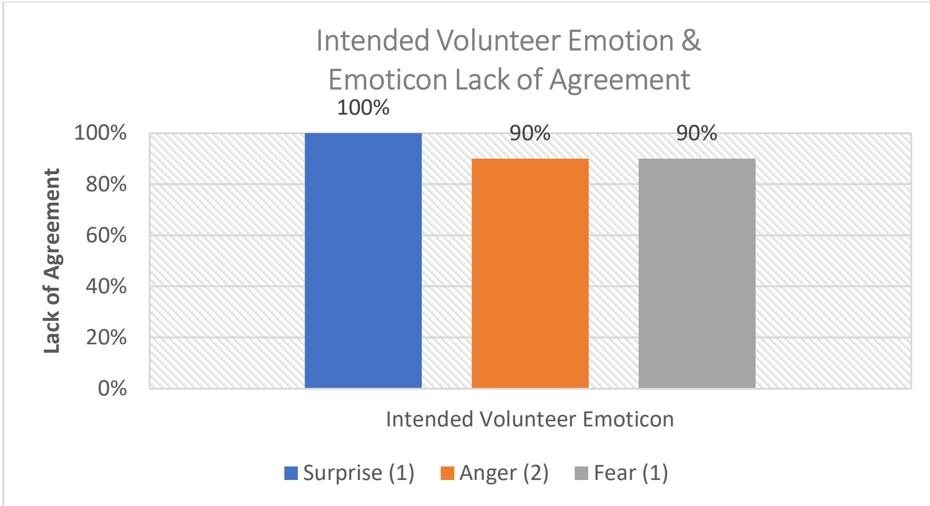


Figure 14. Intended Volunteer Emotion & Emoticon Lack of Agreement

III. Discussion

The current pre-pilot study was the first attempt in the development of the mASkeD app to use the mASkeD app in real-time. The mASkeD app was designed to be used as an intervention tool for people with ASD as part of speech-language therapy with an SLP. The mASkeD app is a low cost, stand-alone app that can be used in various lighting conditions from different angles to identify the emotion of the conversation partner. Before using the mASkeD app with people with ASD, we investigated neurotypically developing people's level of agreement with the emoticon generated by the mASkeD app.

As a first step, participants completed a survey with screenshots of volunteers demonstrating emotions in real-time and the emoticon generated by the mASkeD app. The stimulus photos consisted of a screenshot of the conversation partner with the emoticon generated by the mASkeD app. Since this was the first time the mASkeD app was used in real-time with video streaming, the results revealed many points of consideration as the mASkeD app continues in the developmental phase. The lack of agreement between the volunteers' intended emotions and the emoticons generated by the mASkeD app was initially surprising; however, upon closer consideration, there are several possible explanations. Considering human accuracy may explain some of the variations that were observed in the results. Additionally, the training process that the mASkeD app underwent is possibly a contributing factor to some of the lack of agreement that was observed.

Intended Volunteer Emotions

During stimuli development, there were emotions that the volunteers intended to portray. When the mASkeD app generated an emoticon, the emoticon did not always match the intended emotion of the volunteer. Within the 9 emoticons that matched volunteers' intended emotions, sad (4) was identified with the most accuracy followed by happy (3), fear (1), and neutral (1). The overall agreement across participants for these 9 stimulus pictures was 75.5%. The percent of agreement across participants for stimulus pictures containing agreement with the intended volunteer and emoticon generated by the mASkeD app meets the 75% accuracy rate that the mASkeD app was reported to achieve during previous testing.

Twenty-five (73.5%) of the 34 emoticons lacked agreement between the intended volunteer emotion and the emoticon generated by the mASkeD app. With 70-100% agreement across participants, 4 screenshots did not match the volunteer's intended emotion. The emoticons represented in this category were neutral (2) and fear (2). Differing rates of agreement between intended volunteer emotions and the emoticon generated by the app warrants consideration of the mASkeD app's training process. For example, there was 80% agreement between the intended volunteer emotion of sad and the emoticon generated by the mASkeD app, but only 14.2% agreement between the intended volunteer emotion of fear and the emoticon generated by the mASkeD app.

The mASkeD app never generated the disgust emoticon for any of the volunteers, even though it was an intended emotion. The lack of the generation of the disgust emoticon requires us to consider the role that the training of the mASkeD app may have in the lack of the disgust emoticon.

Training of the mASkeD app. The four emoticons with the highest amount of training images (happy, neutral, sad, and fear) had 100% agreement between people and the emoticon generated by the mASkeD app (see Figure 8). While emoticons with less training images (anger and surprise) had 50% agreement or lower (see Appendix B). Specifically, the emoticon for anger was generated once and there was 50% agreement between people and the emoticon. The emoticon for surprise was generated twice: one stimulus item had 30% agreement and the other stimulus item had 10% agreement between people and the emoticon generated. In total, disgust had the least amount of training images, which could partially explain the emoticon for disgust never being generated.

One additional point to consider with the way in which the mASkeD app was trained is the emotions on which there was agreement among participants regarding the emoticon generated by the mASkeD app, but a lack of agreement with the intended emotion of the volunteer. These emotions included angry, surprise, and fear. With emotions like angry, surprise, and fear there may be some motivation on the part of the conversational partner (volunteer) to downplay these emotions. For example, it may not be socially acceptable to get angry at work. As a result, the way that a volunteer demonstrates angry may be more subtle than the items used during training for angry from the FER-2013 and KDEF datasets. The subtly may cause participants to identify the emotion differently from the intent of the volunteer, a typical occurrence in social interactions with people.

Human Accuracy. The mASkeD app's training focused on identifying emotions as humans do. Human accuracy with the FER-2013 dataset is important to consider. In a small-scale study, humans identified the emotions in the FER-2013 dataset with 65% $\pm$ 5 accuracy (Goodfellow et al., 2013). While this may seem like a low level of accuracy, one must consider all that goes into the identification of emotions. Even individuals that are neurotypically developing do not identify emotions with 100% accuracy. Evaluating the agreement between people and the emoticon generated by the mASkeD app, numbers such as 70% (fear) and 55% (happiness) are not too far off from the accuracy with which humans identified images from the training dataset (Goodfellow et al., 2013).

Considering the role of intended emotions and how emotions are interpreted can inform realistic levels of accuracy to expect. Differences between how the mASkeD app was trained and the way in which it works in real time might partially explain the lower levels of agreement. Interpreting the results from this perspective, evaluating the current limitations of the application of the mASkeD app are warranted to continue with the development of the mASkeD app.

Limitations

Since this was the first time the mASkeD app was evaluated for its use in real-time, some factors are likely contributing to the results. The number of participants was small. However, we wanted to pilot the survey to determine the potential challenges that could occur in larger scale surveys in the future.

Homogeneity of Groups. The volunteers in this pre-pilot study were all related. As a result, all volunteers were from similar ethnic backgrounds. The coronavirus (COVID-19) pandemic limited our access to a more diverse group of volunteers. The requirement for sideloading the mASkeD app onto an Apple device limited the number of people for whom we could install the mASkeD app. This in turn, limited access to volunteers to demonstrate the emotions. With a more heterogeneous group of volunteers in the future, we will aim to have a balanced number of screenshots from each volunteer representing each of the seven universal emotions.

In the current pre-pilot study, 90% of the participants were females from similar ethnic backgrounds. Since the current pre-pilot study utilized a convenience sample, the lack of diversity in the participants will require us to seek out other avenues for participant recruitment in the future. Recruiting participants from different ethnic backgrounds and genders with a variety of educational backgrounds will allow for greater generalizability of results.

Lack of variation in emoticons generated the mASkeD app. There is a possibility that the lack of diversity in volunteers may contribute to the imbalance of emoticons generated by the mASkeD app. During the pre-pilot study, the sad emoticon was generated by the app 13 times, while the angry emoticon was only generated once. Future studies need to consider whether this imbalance is due to a heterogeneous group of volunteers or if there are other issues that need to be addressed. The lack of variation in emoticons generated may also reflect training adjustments that need to be made within the mASkeD app before it is piloted in the future.

The inclusion of photos of actors in the KDEF data set may contribute to the lack of variety in emoticons generated as well. Actors demonstrating emotions for the KDEF dataset were given specific instructions on how to demonstrate the emotions. Perhaps the emotions demonstrated by the actors are not the best training items to use when considering how the emotions are demonstrated in real-life. Adding additional training items from more non-actors might be warranted if it will resolve the problem of the disgust emoticon not being generated during the use of the mASkeD app.

Lack of proportionate training data for each emotion. The mASkeD app was trained with an uneven number of images. This may have contributed to the reduced variety of emoticons generated by the mASkeD app. Refining the mASkeD app by including more training images for disgust is possibly needed. Including more training images for disgust could aid in ruling out the possibility that the lack of training photos is the reason for specific emoticons being generated less often. Also, including a wider variety of training photos for emotions that can be expressed more subtly man increase the accuracy of emotions where there was agreement across participants, even though the *intended* emotion of the volunteer was different.

Future Directions. This pre-pilot study has provided valuable information for the larger study of creating an app that can identify emotions as a tool for individuals with ASD to use as part of therapy with an SLP. Future studies should include survey questions that require participants to identify the volunteers' emotions independently (without the emoticon generated by the mASkeD app). This independent generation of the label for the emotion will help to determine true agreement between people and the mASkeD app. Future studies should include more diverse age groups and ethnic backgrounds in both the volunteer and participant groups. Additionally, studies should be completed to assess if the inclusion of verbal and gestural streaming can increase the amount of agreement between people and the emoticon generated by the mASkeD app.

APPENDIX

Appendix A. Email Recruitment

Recruitment Email Message Template

To: [a_s911@txstate.edu]
From: [a_s911@txstate.edu]
BCC: [Potential Participants]
Subject: Research Participation Invitation: [Facial Expression Recognition App Accuracy]

This email message is an approved request for participation in research that has been approved by the Texas State Institutional Review Board (IRB).

[Dear CDIS Students],

My name is Anna Stewart and I am an undergraduate student at Texas State University. I am conducting a study to evaluate the agreement in the identification of emotions by an app designed for emotion identification and people with no known neurological diagnosis. The information gathered will be used to assess the accuracy of an emotion recognition app**.**

You are being asked to complete this survey because you are a person with no known neurological diagnosis.

If you agree to participate in this study, please click on the survey link provided in this e-mail. This link will open a survey via Qualtrics. The survey will take approximately 60 minutes or less to complete. Your survey answers are completely anonymous. You must be at least 18 years old and have no known neurological diagnosis to take this survey.

Participation is voluntary. You do not have to participate in this study if you do not want to. We ask that you try to answer all questions; however, if there are any items that make you uncomfortable or that you would prefer to skip, please leave the answer blank. If you volunteer to be in this study, you may withdraw from it at any time without consequences of any kind.

Possible benefits from this study include aiding in the development of an application that can be used to identify emotions.

Participants who complete the survey will receive a $10 Amazon e-gift card.

To participate in this research, click the link below.

To ask questions about this research please contact:

Anna Stewart, at a_s911@txstate.edu or

My faculty advisor Dr. Maria Resendiz at mr54@txstate.edu.

Below is the link for the survey:

https://txstate.co1.qualtrics.com/jfe/form/SV_0Hx8kjK5HURSdMN

This project #7176 was approved by the Texas State IRB on 4/1/2020. Pertinent questions or concerns about the research, research participants' rights, and/or research-related injuries to participants should be directed to the IRB chair, Dr. Denise Gobert 512-716-2652 – (dgobert@txstate.edu) or to Monica Gonzales, IRB Regulatory Manager 512-245-2334 - (meg201@txstate.edu).

Pertinent questions or concerns about the research, research participants' rights, and/or research-related injuries to participants should be directed to the IRB

Appendix B. 60% Agreement and Below

Disgust

60 %Agreement
Across
Participants

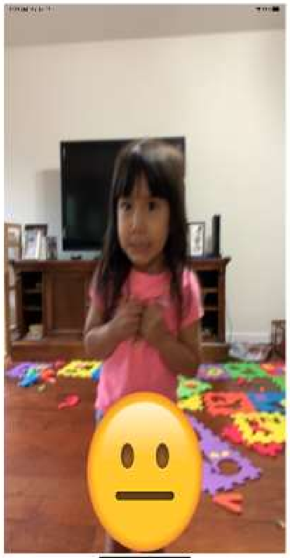
40

Disgust

Sad

50 % Agreement
Across
Participants

41



Fear

Sad

40% Agreement Across Participants

42

Fear

Sad*

30% Agreement Across Participants

43

Fear Neutral Sad Disgust Happy* Surprise

10% Agreement Across Participants

44

REFERENCES

Affectiva. (2017, Oct. 31). Emotion al 101: All about emotion detection and Affectiva' s emotion metrics. *Affectiva.* https://blog.affectiva.com/emotion-ai-101-all-about-emotion-detection-and-affectivas-emotion-metrics

Affectiva. (2020). Deep Learning. https://www.affectiva.com/how/deep-learning-at-affectiva/

Affectiva. (2020). In lab biometric solution. *Affectiva*. https://www.affectiva.com/product/individual-product-page-imotions/

Affectiva. (2020). *Solutions.* https://www.affectiva.com/what/products/

Agrawal, R.K. (29 Mar 2018). *Difference between Machine Learning, Deep Learning and Artificial Intelligence.* Medium. https://medium.com/@UdacityINDIA/difference-between- machine-learning-deep-learning-and-artificial-intelligence-e9073d43a4c3

American Psychiatric Association (2013*). Diagnostic and statistical manual of mental disorders.* 5th ed. Arlington, VA: American Psychiatric Association.

ASHA. *(2020). Role of SLP in Diagnosis.* ASHA. *https://www.asha.org/PRPSpecificTopic.aspx?folderid=8589935303&section=Roles_and_Responsibilities*

Burgoon, J. K., Guerrero, L. K., & Floyd, K. (2016). *Nonverbal communication.* Routledge.

Challenges in representation learning: Facial expression recognition challenge. (n.d.). https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge

Ekman, P., & Friesen, W. V. (1986). A new pan-cultural facial expression of emotion. Motivation and emotion, *10*(2), 159-168. https://doi.org/10.1007/BF00992253

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124.

Frith, C. (2009). Role of facial expressions in social interactions. Philosophical Transactions of the Royal Society B: Biological Sciences, *364* (1535), 3453-3458.

Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., & Zhou, Y. (2013, November). *Challenges in representation learning: A report on three machine learning contests.* In International Conference on Neural Information Processing, 117-124. Springer: Berlin, Heidelberg.

Google AR & VR. (2019, May 20). Glass Enterprise Edition 2: A hands-free device for smarter and faster hands-on work. [Video]. YouTube. https://www.youtube.com/watch?v=5IK-zU51MU4

Google. (2012, Apr 4). Project Glass: One Day [Video]. YouTube. https://www.youtube.com/watchtime_continue=142&v=9c6W4CCU9M4&feature=emb_logo

Haselton, T. (2019, May 20). Google unveils new $999 smart glasses for businesses, undercutting Microsoft's HoloLens on price [Blog]. *CNBC.* *https://www.cnbc.com/2019/05/20/google-glass-enterprise-edition-2-announced-price.html*

Haque, M. I. U. (2019). A facial expression recognition application development using deep convolutional neural network for children with autism spectrum disorder to help identify human emotions.

https://digital.library.txstate.edu/handle/10877/8457

Kaggle. (2013). Challenges in Representation Learning: Facial Expression Recognition Challenge. *Private Leaderboard.* https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/leaderboard

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous child*, *2*(3), 217-250.

Kok, J. N., Boers, E. J., Kosters, W. A., Van der Putten, P., & Poel, M. (2009). Artificial intelligence: definition, trends, techniques, and cases. *Artificial intelligence*, 1.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). About KDEF. *KDEF & AKDEF*. https://www.kdef.se/home/aboutKDEF.html

Marsland, S. (2015). Machine learning: an algorithmic perspective. CRC press.

MATLAB. (2018, Oct. 3). Deep learning: Neural networks basics [Video]. YouTube. https://www.youtube.com/watch?v=437dvcP-09E&feature=share

Perry, A. & Factor, D.C. (1989). Psychometric validity and clinical usefulness of the vineland adaptive behavior scales and the AAMD adaptive behavior scale for an autistic sample1. *Journal of Autism and Developmental Disorders, 19* (1), 41-55. https://linkspringercom.libproxy.txstate.edu/content/pdf/10.1007%2FBF02212717.pdf

Saha, S. (2018). A comprehensive guide to convolutional neural networks—the ELI5 way. *Towards Data Science*. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Stanford University. (2020). About. Stanford Medicine. https://guesswhat.stanford.edu/

Stanford University. (2019). Guess What? Overview. *The Wall Lab at Stanford*

University. https://wall-lab.stanford.edu/projects/guess-what/

Stanford University. (2019). About the Project. *The Autism Glass Project at Stanford Medicine.* http://autismglass.stanford.edu

Stanford University. (2020). *Convolutional Neural Networks for Visual Recognition.* CS231n. https://cs231n.github.io/

Uljarevic, M., & Hamilton, A. (2013). Recognition of emotions in autism: a formal meta-analysis. *Journal of autism and developmental disorders, 43*(7), 1517-1526.

Young-Browne, G., Rosenfeld, H. M., & Horowitz, F. D. (1977). Infant discrimination of facial expressions. *Child Development*, 555-562.

Zijderveld, G. (2017, Apr 14). The World's Largest Emotion Database: 5.3 Million Faces and Counting. *Affectiva.* https://blog.affectiva.com/the-worlds-largest-emotion-database-5.3-million-faces-and-counting