Developing an eBook-Integrated High-Fidelity Mobile App Prototype for Promoting Child Motor Skills and Taxonomically Assessing Children's Emotional Responses Using Face and Sound Topology

William Brown III, DrPH, MA[1,2], Connie Liu, MA[1], Rita Marie John, CPNP, DNP, EdD[3], Phoebe Ford[4]

[1]Department of Biomedical Informatics, Columbia University, New York, NY; [2]HIV Center for Clinical and Behavioral Studies, New York Psychiatric Institute and Columbia University, New York, NY; [3]School of Nursing, Columbia University, New York, NY; [4]School of International and Public Affairs, Columbia University, New York, NY

Abstract

Developing gross and fine motor skills and expressing complex emotion is critical for child development. We introduce "StorySense", an eBook-integrated mobile app prototype that can sense face and sound topologies and identify movement and expression to promote children's motor skills and emotional developmental. Currently, most interactive eBooks on mobile devices only leverage "low-motor" interaction (i.e. tapping or swiping). Our app senses a greater breath of motion (e.g. clapping, snapping, and face tracking), and dynamically alters the storyline according to physical responses in ways that encourage the performance of predetermined motor skills ideal for a child's gross and fine motor development. In addition, our app can capture changes in facial topology, which can later be mapped using the Facial Action Coding System (FACS) for later interpretation of emotion. StorySense expands the human computer interaction vocabulary for mobile devices. Potential clinical applications include child development, physical therapy, and autism.

Introduction

The importance of stories in physical and emotional development, and health practice, cannot be overstated. Stories are a universal concept across cultures that predate the written word. Stories have been told from prehistoric to contemporary to modern times. They also transcend multiple cultural barriers: Language, age, culture, and education. As a result, many stories are shared across cultures and/or are retold with differentiating degrees of variation and purpose. The earliest of storytelling was oral and heavily combined with gestures and expression to enhance the effectiveness of conveyance.[1] As a result, stories can also be a mechanism for physical development and have the ability to elicit emotional responses. Unfortunately, gesture and expression in storytelling has not translated to mobile technology (i.e. eBooks) to the same degree.

Computers and information technology are relatively new to the process of storytelling. Newer still are mobile devices. However, their growing ubiquity and advancing technology potentiate storytelling in new, dynamic, and undiscovered ways. Commonly, mobile devices rely on screen input interface and touch to simulate the standard functions of books. Swiping your finger from one side of the screen to the other performs page turning, and page leaves are animated to resemble real pages. Tapping and pinching are additional features offered by mobile devices. They allow zooming into pages for better views and selection of text to capture content for alternate uses.

Unfortunately, these actions do not meet the range of motion necessary for a child's comprehensive "motor skills" development. They also do not provide sensory and algorithmic identification methods for confirming that the young reader, or the child being read to, has actually performed the task. However, there are many unused mobile device sensors that, if leveraged, can: 1) enhance the way children use motor skills while experiencing eBooks, 2) confirm the execution of specific motor skills, and 3) provide a better understanding of the child's emotional experience during the course of their interactive eBook experience.

We identified several untapped mobile device sensor resources ideal for identifying motor development and assessing emotion during storytelling when using eBooks. We chose to focus on the two sensors that are least used, yet ubiquitously available on most mobile devices. The two sensors are the camera for image capturing and the microphone to capture sound. Advances in camera technologies more accurately simulate the capability of an eye where small and large ranges of motion can be captured, as well as objects detected and identified. Sound detection

is the most common feature of any mobile device, and is able to detect and differentiate greater ranges of sound than ever possible.

Given the knowledge gap related to mobile device assisted motor development and emotion detection, and the opportunities provided by new mobile technologies, we aimed to answer the question "Can an ebook-integrated high-fidelity mobile app prototype be developed for promoting child motor skills and taxonomically assessing children's emotional responses?" The purpose of this work was a proof-of-concept to develop "StorySense", an app that can promote child motor skills and identify emotion, for healthy child development. Our development goal included creation and testing of two sensory functions and identifying necessary programming and topology classification libraries. To evaluate our work we employed a system development life cycle stage-based evaluation model.

Theoretical base

The theoretical basis for this application's use in child development comes from the fundamentals of interactionist theories such as the neuronal group selection theory and the dynamic system theory.[2,3] These theories support that the infant's motor abilities emerge as a result of the child, task, and the environment with individual variability.[4] The plasticity of the brain in children is the neurophysiological basis for promoting motor skills in young children. Research over the past fifty years has confirmed that an infant's brain is built over time and that the development allows for future skills to emerge. Brains are modulated by genetics and experiences that will affect the outcome of the child.[5] StorySense provides interactive experiences similar to those that are known to have a modulation effect on the outcome of a child's gross and fine motor development.

Expression through facial action, gesticulation, and sound

Facial expressions are fundamental to the communication of simple and complex emotion. Movements in the muscles and skin, particularly around the mouth and eyebrows, provide a large visual vocabulary of meaning and emotional terminology.[6] Similarly, gesticulations with hands and the creation of sound (e.g. clapping, rubbing, banging, tapping, snapping) are intrinsically tied to communicating, and can add depth to conveying emotional information while children experience eBooks.[7]

There are two types of facial expressions that contribute to communication, voluntary expression and emotional expression.[8] Voluntary expression follow learned display rules in emotion and are made consciously (e.g. blowing a kiss). Emotional expression, on the other hand, is often displayed unconsciously. This includes facial expressions like distress, disgust, interest, anger, contempt, surprise, and fear. Despite the fact that individuals do not realize they are producing these expressions, this visual autonomic vocabulary is information rich and universally comprehendible.[6] Thus, the eyes and mouth are a fundamental identifier in facial recognition.

Sounds are also used for processes of identification. Sound variation and scale carry varied meaning (e.g. clapping, screaming, and whistling). It is an ideal indication of information that is both consciously and subconsciously conveyed by potential technology users. Both sound and facial movement reveal feelings and thoughts. However, the range and types of feelings and thoughts can be very different and very unique to one mode of information conveyance versus the other. While facial movement can indicate things like attraction, disgust, uncertainty, sound on the other hand can indicate specific emotions through utterances, speech, and various onomatopoeia (e.g. hissing).[7] StorySense both leverages and builds on previous research in: face, motion, and sound recognition; facial action coding and emotion identification; sensing for people-centric applications; and reaction sensing. Below, we discuss work relevant and contributory to this project.

Facial Action Coding System (FACS)

The systematic categorization of facial movements to identify expression of emotion is a historic practice of psychologists. In 1978 Paul Ekman and W.V. Friesen developed a Facial Action Coding System (FACS) by analyzing changes in facial appearance created during various combinations of facial muscular contraction. Their goal was to develop a reliable scoring metric by which human raters can identify facial behaviors. The result was FACS a taxonomic system of human facial movements that can help raters code changes in elements of the face (i.e. eyes, mouth) and their muscular movements.[9]

Today, FACS is the most widely used descriptive measurement tool for facial behaviors, and aids computers to topologically detect faces and their geometry. It also allows for the reduction of subjectivity and the use of high-throughput computational methods. FACS measurement units are Action Units.[9] Consequently, FACS scores do not provide meaning of facial behavior, and can only be used descriptively. To address the need for meaningful interpretation of FACS scores, researchers developed the Facial Action Coding System Affect Interpretation Dictionary (FACSAID). FACSAID links facial expressions with their psychological interpretations and models them to facial behaviors, then stores this information in to a relational database.[10] Thus, the raw FACS scores potentially produced by a StorySense scan can be translated into more psychologically meaningful concepts. By leveraging both the FACS taxonomy and FACSAID as knowledge bases, the final version of StorySense will be able to produce emotional profiles of a reader's facial movement for clinical interpretation and possible therapeutic use.

Sensing for people-centered mobile applications

There have been many applications of sensing technology for people-centered mobile applications. Cameras are one of the most utilized features in mobile devices and have some of the most varied function. Originally the camera's function was solely relegated to pixilated single pictures. Now we find high resolution images that can be instantly manipulated, as well as motion cinematography.[11] Similarly, the microphone has advanced beyond the standard receiver and plays a larger role in capturing sound. It pairs with camera recordings to produce video recordings, takes dictation, and is used to input commands.[12] As previously mentioned accelerometers are at the forefront of much of the motion detection capabilities of mobile devices.[13] What's more, when paired with GPS systems and the gyroscope, the duo maximize the devices ability to orient itself in the universe.[12–14] Though this information can also be helpful in detecting the movement of the user, the combination of accelerometers and GPS are most often an indicator of distance traveled. The amount of distance traveled from one physical location to another is likely to be short and of less relevance during a child's experiencing while reading an eBook. Thus, we did not focus on these to available features.

Reaction sensing

A developing area in HCI is reaction sensing. Reaction sensing goes beyond the normal bounds of input and imputation. Researchers and developers are leveraging sounds and their variations in order to enhance HCI and create a multifaceted user experience.[15] Moreover, advanced reaction sensing leverages the behavioral cues that are autonomic to people.[15,16] In this way the interaction more easily mimics what a person would expect another person to react to. For instance, previous work in sound recognition includes scalable sound sensing for people-centric applications on mobile phones (i.e. SoundSense).[15] Work in face detection includes technologies such as blink detection for real-time eye tracking.[16]

Moreover, advances such as natural language processing (NLP), machine learning, and data mining, can combine with multimodal sensing technologies, and allow mobile devices to anticipate user needs and react organically to people's naturally occurring behavioral cues.[16–18] This has the added benefit of being able to track and log new and existing movements and reactions. As a result, the potential to increase the known vocabulary of reactions to stimuli means we can process, analyze, and understand behavioral cues collectively and longitudinally, making high-throughput analyses a real possibility.

Methods

We built StorySense in Java for Android. We successfully tested and ran our applications on a Nexus 7 tablet and Samsung Galaxy S III mobile phone. We focused on two areas of sensing, sensing sound produced by high motility (clapping) and low motility (snapping) processes, as well as detecting facial changes, with a focus on eye movement. We also systematically identified classic children's stories that were ideal for incorporating movement and promoting motor skills.

Criterion and methods for choosing a story for the prototype

We defined story criteria to optimize our application's user experience, diversify options for interaction, and integrate fluidly with new sensing techniques. Our criteria included: English availability, multiple translations (for universality), internationally stable storyline, high graphic artwork, abstract storyline, availability/expired copy write, and animation potential. Using the Google search engine and leveraging its "Scholars" database, we performed a library literature search of thousands of choices. Based on our criteria we narrowed our results down to three possible stories: Peter Pan, Cinderella, and The Wizard of Oz. All researchers reviewed the final three stories. Ultimately, we chose Peter Pan because, of the three story options, it had the best balance of being well known and a stable storyline across cultures. The stable storyline across cultures was the penultimate criteria because it is an ideal trait for future evaluation.

In addition to identifying an optimal story for the prototype, we also had to edit the features of the story and create additional interactive illustrations and communicate functionality (Figure 1).

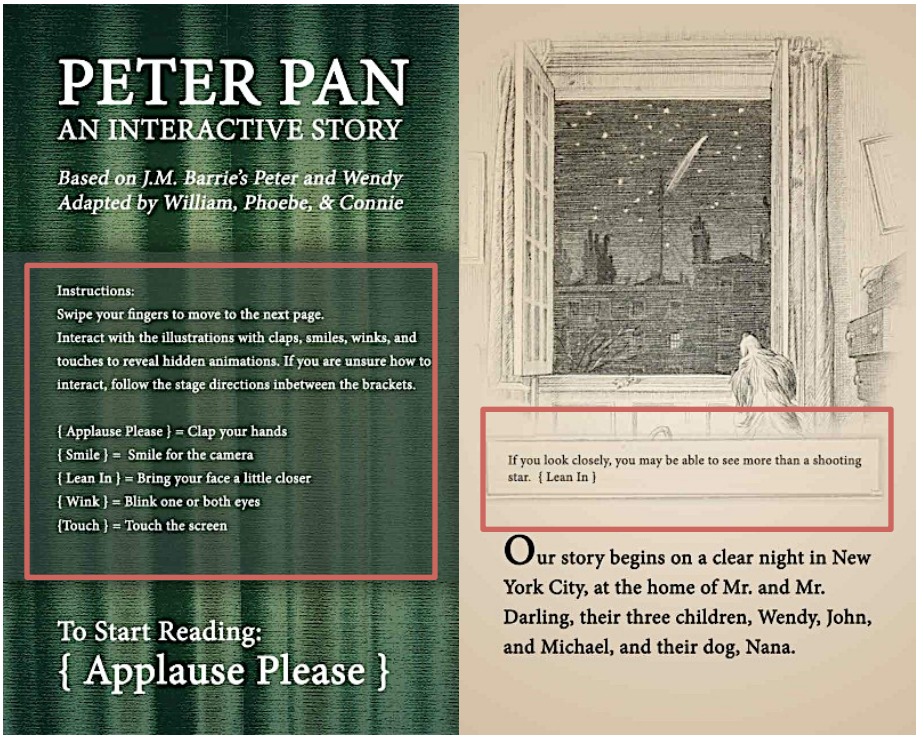

Figure 1. Modified illustrations (outlined in red).

Sensing sound: Clap/snap detection

For the purposes of prototyping, we chose to focus on clap and snap detection. We plan to explore detecting a wider range of sounds in the future. In consideration of the broad range of stories available as eBooks, it would be useful to detect and respond to a wider range of human sounds, such as sneezes, laughs, and snorts. However, sounds such as those we just listed are not produced using high-motility actions. Future noises that could be explored beyond clapping that are produced through high-motility actions include stomping and banging. However, we were unable to find these sound types to be useful in the same story. Originally we focused only on clap detection, however we needed a low-motility sound and verbalization (i.e. talking) for performance juxtaposition and sound identification. Thus, we quickly incorporated snap detection as well. This serves two purposes. In conditions where clapping is the preferred action; we want to make sure that sounds can be distinguished from similar sound patterns. This way the eBook can ensure that the desired physical activity is being performed. Secondly, we want to accommodate users who prefer to hold a mobile device with one hand and use snap detection. Furthermore, a story may elicit these sounds from its audience and perform a pre-designated response or action in return, or just respond to organically produced sounds as they happen naturally.

Sensing the face: Eye detection

We use facial recognition to detect and track user eye movement. Triggered by a user click, our application accesses the mobile device's camera and initiates a series of actions for facial recognition and eye tracking (Figure 2). In the development of the face recognition portion of the application, several decisions were made in regards to how we access a device's camera and how a user would be interacting with our application. First, we found there to be two different ways to capture images using the camera on Android. One is using an intent, which is using an existing camera application to take the photo. Second, is to create a custom camera application to take the photo. The first option would have another application come into the foreground, thereby occluding the story page and disrupting story flow. Thus, we chose the second option.

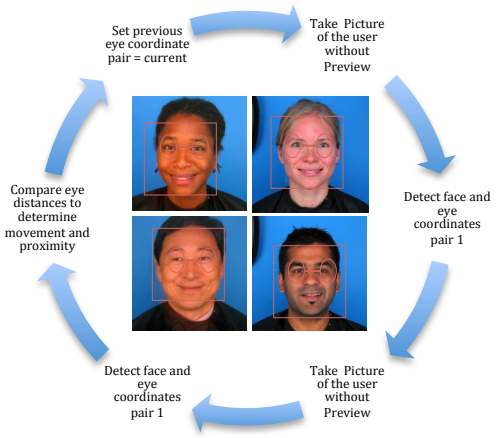

Figure 2. Face and eye detection. (Images from PICS database[19])

However, the second option had the Android required condition where in order to take a picture, there has to be a preview surface, meaning the user has to see what they are taking a picture of and then click to take a picture. The problem here is that we did not want a preview since it would occlude much of the story page as well as disrupt story flow. Ultimately, in order to get around this issue, we created a surface view and resized it to a single pixel on the screen, thus there is a preview, but it is not visible.

Evaluation

Development and evaluation of StorySense followed the systems development life cycle (SDLC), also referred to as the application development life-cycle, we have completed the preliminary analysis, systems analysis, requirements definition, systems design, development phases, integration and testing phase.[20] This evaluation highlights our experience and knowledge gained in the integration and testing phase. The following evaluation results are based on StorySense's current SDLC stage, benchmark lab testing.

Benchmark lab evaluation of system sensors

To evaluate StorySense's ability to detect and distinguish sound topology (i.e. talking, clapping, snapping) we used threshold detection. We looked at amplitude readings from mobile device and laptop microphones for a variety of sounds to determine a threshold. We observed amplitudes for: clapping slow and fast; clapping close and far; clapping slow and fast while talking; snapping; talking; yelling; and tapping the device. We ran these tests with two users, one who snaps loudly and one who snaps quietly. These performance tests were done while watching an amplitude meter to help us understand the basic differences in the amplitude over time. We recorded our results in the system and mapped clapping topologies to event triggers in the story.

We evaluated StorySense's ability to detect eye topology using visual tracking confirmation methods. The series of actions following the user click included capturing the user's photo using the front facing camera without a preview, thereby not disturbing or occluding the story page. The picture taken is then sent to a built-in face detector, and if a face is detected, the eye position is recorded. With the eye screen coordinates in-hand, a Peter Pan character image is drawn on the story page screen where the eyes were detected (Figure 3). That is, once we have one pair of eye coordinates; our camera module is able to take another user photo and extract the eye coordinates from that. If at least two pairs of eye coordinates are available, the application draws a Peter Pan character image at both eye locations and a dashed line between them to show the user's eye movement was tracked. This Peter Pan image allowed us to detect changes in eye topology, and was used in the prototype model for verification of metrics, direction, and actions performed.

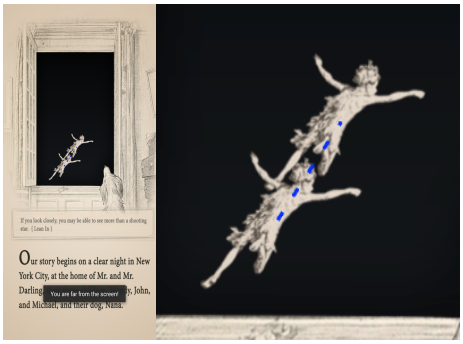

Figure 3. Eye tracking/Peter Pan indicators.

Furthermore, we conducted a performance evaluation of StorySense's face detection software, which consisted of testing if the application is able to detect faces (including recording the time used for detection in milliseconds [converted to seconds], and the resulting confidence factor) for a subset of 40 images from the Psychological Image Collection at Stirling (PICS) University database.[19] The image set from PICS included both ethnic and phenotypic diversity, and consisted of 22 male and 18 female images. The confidence factor used to gauge face detection is a property of the Android Software Development Kit's (SDK) "FaceDetector." This property is also known as the "Face Class", which holds information regarding the identification of a face in a bitmap. The confidence factor returns a value between 0 and 1, and it indicates how certain what has been found is actually a face.

Results

Sound detection evaluation results

The graphs below show that claps and snaps have significantly higher amplitudes than "indoor voice" talking at close range (~ a foot and a half away) (Figure 4). Another difference is that the claps and snap amplitudes go back down to zero very quickly, while talking has more of a zigzagged amplitude corona arch. Both of these features can be combined to yield reasonable results for clap and snap detection. While currently StorySense identifies anything above a threshold as a clap or a snap, it could benefit from further analyzing the sound following the initial trigger to determine if the sound is staying high or going back down quickly. This added feature would dramatically increase the ability of StorySense to detect claps and snaps.

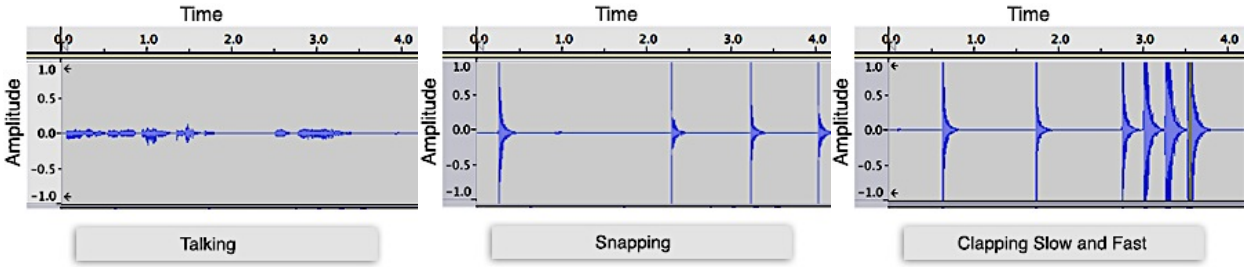

Figure 4. Sound types and amplitudes.

Unfortunately, the threshold we chose to use as our base level made it difficult to distinguish additional sounds such as snaps or short bursts from claps. If someone yells loudly or a nearby dog barks, the sounds will be registered as a clap and the system will respond accordingly. In general, people usually talk with "indoor voices" that stay below the threshold or read silently. More formal user testing could help to define the accuracy of this approach.

Face detection evaluation results

We were able to successfully procure eye coordinates; we also obtained the distance between the eyes. We were then able to use this information to determine additional changes in facial topology as well as the distance of the user from the screen. We could accurately calculate distance of user using eye size. The larger the eye was, the closer the user's face was to the screen. We could also use these variations in distance and eye size to detect eye blinking, proximity, and various movement types. Alternatively, if the user was too close (i.e. distance is greater than mobile device screen size threshold), then a warning successfully triggered and posted to the screen, telling the user to increase their distance from the face of the mobile device. This would remind users not to put their faces too close to the screen; thus, maintaining a safe viewing distance, as well as facial detection accuracy.

According to the Android application program interface (API) documentation, a confidence factor above 0.3 indicates good face detection. Our results from our face detection performance evaluation (Table 1) show that we have a 100% face detection (FD) rate, with average detection time being 2772 milliseconds (or 2.772 seconds) and average confidence factor being 0.522917008.

Table 1. Face detection performance evaluation (Time for face detection (FD) was converted from the original output in milliseconds to seconds).

	Image	FD?	Confidence Factor	Time for detection (Seconds)		Image	FD?	Confidence Factor	Time for detection (Seconds)
1	f4001s.jpg	Y	0.5312345	2.747	21	m4003s.jpg	Y	0.51446885	2.911
2	f4002s.jpg	Y	0.5130435	2.567	22	m4004s.jpg	Y	0.52985317	2.989
3	f4003s.jpg	Y	0.5289631	2.476	23	m4010s.jpg	Y	0.5295981	3.015
4	f4004s.jpg	Y	0.5270577	2.572	24	m4011s.jpg	Y	0.5311005	2.721
5	f4005s.jpg	Y	0.51732284	2.555	25	m4012s.jpg	Y	0.5211092	2.748
6	f4006s.jpg	Y	0.52806115	2.43	26	m4014s.jpg	Y	0.5209863	2.774
7	f4007s.jpg	Y	0.5279383	2.833	27	m4017s.jpg	Y	0.51728237	2.822
8	f4008s.jpg	Y	0.5243423	2.829	28	m4018s.jpg	Y	0.52777547	2.931
9	f4009s.jpg	Y	0.525947	2.888	29	m4021s.jpg	Y	0.5211933	2.963
10	f4010s.jpg	Y	0.5292073	2.837	30	m4024s.jpg	Y	0.5209574	3.002
11	f4016s.jpg	Y	0.512235	3.319	31	m4027s.jpg	Y	0.5180155	2.487
12	f4017s.jpg	Y	0.52486527	2.984	32	m4028s.jpg	Y	0.5112237	2.448
13	f4018s.jpg	Y	0.53505766	3.539	33	m4031s.jpg	Y	0.51334447	2.452
14	f4021s.jpg	Y	0.5200643	2.963	34	m4032s.jpg	Y	0.52210337	2.396
15	f4026s.jpg	Y	0.5255496	3.149	35	m4035s.jpg	Y	0.5270095	2.349
16	f4027s.jpg	Y	0.5266327	2.86	36	m4037s.jpg	Y	0.5241929	2.455
17	f4029s.jpg	Y	0.5280881	2.851	37	m4040s.jpg	Y	0.5187667	2.527
18	f4030s.jpg	Y	0.52181965	2.869	38	m4043s.jpg	Y	0.53127307	2.477
19	m4001s.jpg	Y	0.51076734	2.992	39	m4063s.jpg	Y	0.51823	2.47
20	m4002s.jpg	Y	0.5165685	2.889	40	m4064s.jpg	Y	0.52343065	2.794

Images were taken from the Psychological Image Collection at Stirling (PICS) University database: Category = 2D Face Sets (http://pics.psych.stir.ac.uk/2D_face_sets.htm); Set Name = Utrecht ECVP; Set Description = 131 images, 49 men, 20 women, collected at the European Conference on Visual Perception in Utrecht, 2008. Some more to come, and 3d versions of these images in preparation; Resolution = 900x1200 color.

Once eye movement and distance have been detected and calculated this data can be used to decompose the movement into specific AUs and extrapolated to produce a FACS score. Further, duration, intensity, and asymmetry can improve the accuracy of the score and its related FACSAID interpretation. Although we were currently only tested two uses of eye movement detection, the tracking capability provides possibilities for detecting and identifying other changes in face topology.

Discussion

The stage of our project is proof-of-concept. Our app currently detects discrete interactions, and contains system parameters to detect faces and sound with limited variation for both. Only a sample of the story is being used. System statistics indicate we have reason for concern that there will be device memory space issues when the full story is developed and running; thus, we have focused much of our subsequent work around optimizing the framework to ensure that it handles the loading of pages and animations more efficiently.

Better classifiers to differentiate between sets of sounds and images are highly necessary. Further, development of a larger sound and visual vocabulary and repository are additional goals for the prototype. We may be able to create a

foundational list of desired motor functions from our user survey and Morae 3.3 assessment tool during our first user evaluation, and then compare this list to the existing motions that can be detected by other apps in the Google play store library. Lastly, an immediate goal is to use OpenCV to improve image processing and facial feature detection.

What we have learned in the developmental phase of our app prototype are some of the optimal methods and barriers to creating an app that can be used to detect physical activity of children during their reading of an eBook. We have also identified ways that such functionality can be incorporated directly with the story. In the story of "Peter Pan", the section that tells the reader to clap their hands to bring Tinkerbelle back to life is a clear example of stories moments that directly elicit physical action to produce a story outcome.

Potential clinical applications and target populations

There are no studies using these kinds of application to promote childhood development. It has been shown that early intervention programs that focus on promoting motor skills, preverbal skills, and stimulation of brain development are highly effective. These programs use a variety of techniques that involve multiple senses. The cumulative experience with early intervention programs has confirmed these methods are effective in promoting infant development.[21] StorySense encourages children's motor skills and provides impetus for a response. The functions of StorySense are ideal for increasing child motor skills, and understanding the child's experience even when the child's ability to orally communicate is un-developed.

Our application could potentially facilitate communication skills and social interaction, as well as sustain the child's attention. For children and adolescents with autistic spectrum disorder (ASD) presenting with limitations in conventional forms of verbal and non-verbal communication, this application could provide an alternative form of therapy. Dependent upon the results of our future user evaluation, we could explore the realm of our application as an ASD clinical therapy that could identify limitations and weaknesses in children, as well as strengths and potentials.[22] Furthermore, we could explore application effects on communicative behavior, language development, emotional responsiveness, attention span and behavioral control over a period of time. It may also be possible to put real-time user specific health or physical activity information directly in the eBook story line to help both parents that are reading to their children and children readers to learn how to manage their health.

Contribution to the field(s)

StorySense contributes to the fields of facial recognition, expression coding, and dynamic learning by following the FACS standard of classification protocols and adding to its lexical library. Continued development of StorySense facial recognition algorithms should be able to code nearly any anatomically produced facial expression. During the story telling process, and guided by FACS and FACSAID, StorySense will be able to topologically deconstruct the child's facial movements into specific Action Units, calculate a FACS score, and identify the expression's related emotion. The AUs are independent of variations in human interpretation; thus, higher order decision such as the recognition of basal emotions, can be processed. Subsequently, once a facial expression and its related emotion have been identified, related actions can be pre-programmed into the eBooks story environment. Thus, emotion can also have an impact on the story line in a way that can be designed to address distress, anger, or discomfort, as well as leveraged for therapeutic use.

Future work

Our future work will aim to include geo-location data to trigger a function or change in the story line that could promote walking or encourage other movement (e.g. landscape animations could match the reader's location). We will also improve our application experience by incorporating references to points of interest near the user's current location.

Other future work will incorporate newer technologies and mobile device capabilities. For instance, recent developments such as spritzing technologies (http://www.spritzinc.com/blog/) could provide a way to improve a user's reading skills, speed and focus. Additionally, motion-sensing capabilities of mobile devices could open up a whole new channel of facilitating physical development, conceptually similar to what the Wii Remote motion sensing technology for Wii Sports does.

We were unable to build in a response to other high-motility actions (such as stomping or waving) due to lack of classifiers. Thus, we hope to contribute more classifiers to the established lexicons. Future programming will continue to build off of the work done around reaction sensing.[15,16] Part of our work will also include the development of a facial expression-to-emotion library that is common to children during the process of reading, or is able to distinguish facial movements that are specific to the reading process that may be easily confused with expressing emotion. Additional future work may include train a user specific image set to provide better user recognition with each application use. Thus, future application development will look to implement a more seamless image capture and camera process. More standard functionality such as dictionaries, word pronouncers, bookmarks, and highlighting will also be added.

Future evaluation

Future evaluation will involve usability testing using a user-centered iterative design process, and Morae 3.3 a usability data collection tool to capture audio, video, on-screen activity, and keyboard/mouse/touch input so that we may identify use and error patterns and gain insight into the effectiveness and acceptability of the mobile app's design. NVivo will be used to analyze qualitative data to provide insight into areas for future development.

Limitations

In the development of our application, we had difficulties with sound classification. Loud ambient noises may register as constant clapping. Moreover, as we added animations and more images to test the framework, it became apparent that there were memory related performance issues. Additionally, though face movement can be detected, the app does not yet differentiate between open/close eyes or mouths, which hindered wink and smile detection. Again, we plan on developing classifiers that advances the built-in functions of the mobile device to ameliorate these issues. Also, our sound threshold for testing made it difficult to distinguish additional sounds such as snaps or short bursts from claps. Other loud sounds are at risk of being registered as a clap and the system will respond accordingly. Future iterations of the application might incorporate OpenCV (A library of programming functions mainly aimed at real-time computer vision, http://opencv.org/), open source data sets, and/or more advanced libraries to improve overall face and sound detection.

Using a set of captured 2D images versus actual users is a possible limitation to our evaluation. Moreover, we have thus far only tested StorySense on a Nexus 7 tablet and Samsung SGH-T999 Android 4.1.2 (API 16) (Galaxy S3) 1.9 Mega pixel camera, HD recording @30fps with Zero Shutter Lag, BSI. Newer phone models and other devices that are more widely used (e.g. Apple iPhone) would not be immediately supported. We look to remedy this by creating an iOS version and using emulators as well as acquiring additional mobile devices for testing.

Lastly, the FACS taxonomy is based off of analyzing adult facial movements. There may be unforeseen variations in adult facial recognition. Also, the FACSAID database is based off of FACS codes, which may present a similar problem when classifying child emotions using an emotion dictionary database based on analysis from adult populations.

Conclusion

We accomplished a significant amount of our interactive goals. In the development of this prototype we explored sound and facial expressions, which could potentially contribute to several universal knowledge bases for reaction sensing (i.e. FACS). StorySense leverages unconscious and conscious cues in every day communication. We found that by using mobile sensory technology (i.e. camera and microphone), and integrating common algorithms, it is possible to obtain rich user information. In this way, stories on mobile devices can not only be read by the user, but the mobile device can read the user's conscious or unconscious emotional expression and cues, and adjust the story for a more dynamic and user-centered experience. We also found that sensors used for dynamic and user-centered experiences can be leveraged to promote high-motility action, which would aid in the development of motor skills for young children or children with developmental challenges.

Acknowledgments

Dr. William Brown III is supported by NLM research training fellowship T15 LM007079 and NIMH center grant P30 MH43520. We would also like to thank Dr. Suzanne Bakken for her review on an earlier version of the manuscript.

References

1. VanSledright B, Brophy J. Storytelling, Imagination, and Fanciful Elaboration in Children's Historical Reconstructions. Am Educ Res J. 1992 Dec 21;29(4):837–59.
2. Edelman GM. Neural Darwinism: The theory of neuronal group selection. New York, NY, US: Basic Books; 1987. 371 p.
3. Thelen E. The (re)discovery of motor development: Learning new things from an old field. Dev Psychol. 1989;25(6):946–9.
4. Hadders-Algra M. The Neuronal Group Selection Theory: a framework to explain variation in normal motor development. Dev Med Child Neurol. 2000;42(8):566–72.
5. Adams RC, Tapia C, Murphy NA, Norwood KW, Adams RC, Burke RT, et al. Early Intervention, IDEA Part C Services, and the Medical Home: Collaboration for Best Practice and Best Outcomes. Pediatrics. 2013 Oct 1;132(4):e1073–e1088.
6. Ekman P, Friesen WV. Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues. ISHK; 2003. 200 p.
7. Quek F, McNeill D, Bryll R, Duncan S, Ma X, Kirbas C, et al. Multimodal human discourse: gesture and speech. ACM Trans Comput-Hum Interact. 2002;9(3):171–93.
8. Matsumoto D, Ekman P. Facial expression analysis. Scholarpedia. 2008;3(5):4237.
9. Ekman P, Friesen WV. Facial action coding system: A technique for the measurement of facial movement. Palo Alto: CA: Consulting Psychologists Press; 1978.
10. Merten J. Facial microbehavior and the emotional quality of the therapeutic relationship. Psychother Res. 2005;15(3):325–33.
11. Reynolds F. Camera Phones: A Snapshot of Research and Applications. Pervasive Comput IEEE. 2008;7(2):16–9.
12. Howell J, Schechter S. What You See is What they Get: Protecting users from unwanted use of microphones, camera, and other sensors. In Proceedings of Web 20 Security and Privacy Workshop. 2010.
13. Kwapisz JR, Weiss GM, Moore SA. Activity recognition using cell phone accelerometers. SIGKDD Explor Newsl. 2011 Mar;12(2):74–82.
14. Kim D, Kim J, Choa M, Yoo SK. Real-time Ambulance Location Monitoring using GPS and Maps Open API. Conf Proc IEEE Eng Med Biol Soc. 2008;2008:1561–3.
15. Lu H, Pan W, Lane ND, Choudhury T, Campbell AT. SoundSense: scalable sound sensing for people-centric applications on mobile phones. Proceedings of the 7th international conference on Mobile systems, applications, and services [Internet]. New York, NY, USA: ACM; 2009 [cited 2013 May 15]. p. 165–78. Available from: http://doi.acm.org/10.1145/1555816.1555834
16. Li K. Automatic Content Rating via Reaction Sensing. J ACM. 2013;
17. Morris T, Blenkhorn P, Zaidi F. Blink detection for real-time eye tracking. J Netw Comput Appl. 2002 Apr;25(2):129–43.
18. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc. 2011 Sep 1;18(5):544–51.
19. University of Stirling. Psychological Image Collection at Stirling (PICS) [Internet]. Department of Psychology [Online]; Available from: http://pics.psych.stir.ac.uk/
20. O'Brien JA. Management information systems. 10th ed. New York: McGraw-Hill/Irwin; 2011. 673 p.
21. Dreyer BP. Early Childhood Stimulation in the Developing and Developed World: If Not Now, When? Pediatrics. 2011 May 1;127(5):975–7.
22. Srinivasan SM, Bhat AN. A review of "music and movement" therapies for children with autism: embodied interventions for multisystem development. Front Integr Neurosci [Internet]. 2013 Apr 9 [cited 2014 Mar 13];7. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3620584/