A portable audio/video recorder for longitudinal study of child development

Citation: Soroush Vosoughi, Matthew S. Goodwin, Bill Washabaugh, and Deb Roy. 2012. A portable audio/video recorder for longitudinal study of child development. In Proceedings of the 14th ACM international conference on Multimodal interaction (ICMI '12). ACM, New York, NY, USA, 193-200.

As Published: http://dx.doi.org/10.1145/2388676.2388715

Publisher: Association for Computing Machinery (ACM)

Persistent URL: http://hdl.handle.net/1721.1/80375

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

A Portable Audio/Video Recorder for Longitudinal Study of Child Development

Soroush Vosoughi
MIT Media Lab
75 Amherst St, E14-574K
Cambridge, MA 02139
soroush@mit.edu

Matthew S. Goodwin
Northeastern University
312E RB, 360 Huntington Ave
Boston, MA 02115
m.goodwin@neu.edu

Bill Washabaugh
Hypersonic Engineering and Design
33 Flatbush Avenue, 7th Floor
Brooklyn, NY 11217
bill@hypersoniced.com

Deb Roy
MIT Media Lab
75 Amherst St, E14-574G
Cambridge, MA 02139
dkroy@media.mit.edu

ABSTRACT

Collection and analysis of ultra-dense, longitudinal observational data of child behavior in natural, ecologically valid, non-laboratory settings holds significant promise for advancing the understanding of child development and developmental disorders such as autism. To this end, we created the Speechome Recorder - a portable version of the embedded audio/video recording technology originally developed for the Human Speechome Project - to facilitate swift, cost-effective deployment in home environments. Recording child behavior daily in these settings will enable detailed study of developmental trajectories in children from infancy through early childhood, as well as typical and atypical dynamics of communication and social interaction as they evolve over time. Its portability makes possible potentially large-scale comparative study of developmental milestones in both neurotypical and developmentally delayed children. In brief, the Speechome Recorder was designed to reduce cost, complexity, invasiveness and privacy issues associated with naturalistic, longitudinal recordings of child development.

Categories and Subject Descriptors

H.5.2 [**User Interfaces**]: Graphical user interfaces(GUI); H.5.1 [**Multimedia Information Systems**]: miscellaneous; J.4 [**Social and behavioral sciences**]: psychology

General Terms

Human Factors, Design, Reliability

ICMI'12, October 22–26, 2012, Santa Monica, California, USA.

Keywords

multimedia recording, longitudinal recording, naturalistic recording, video annotation, speech annotation, privacy management, video recorder, audio recorder, child development

1. INTRODUCTION

Collection and analysis of ultra-dense, longitudinal observational data of child behavior in natural, ecologically valid, non-laboratory settings holds significant promise for advancing the understanding of both typical and atypical child development. To that end, we developed the *Speechome Recorder (SR)*(Figure 1), a portable version of the Speechome audio/video recording technology developed for the Human Speechome Project (HSP)[22]. Launched in 2005, the goal of the HSP was to study early language development through collection and analysis of longitudinal, ultra-dense and naturalistic recordings daily in the home of a family with a young child. A total of 14 microphones and 11 omni-directional cameras were installed throughout the house. Video was recorded with 1 megapixel resolution at 15 frames per second using cameras with fisheye lenses embedded in the ceiling. Audio was recorded at 16-bit resolution with a 48 KHz sampling rate using ceiling mounted boundary-layer microphones. Due to the unique acoustic properties of the boundary layer microphones, most speech throughout the house was captured with sufficient clarity to enable reliable transcription. Overall 230,000 hours of high fidelity, synchronized audio-video was recorded continuously for the first three years of the child's life[21], wherein approximately 2.5 million utterances have been manually transcribed to date.

The HSP captured one child's development in tremendous depth. While this corpus is illuminating in a myriad of ways, it focuses on one child of typical development, limiting tests of generalizability with other typically developing children and/or children with developmental delays. The original audio/video recording technology used in the HSP had several issues that did not lend itself to installation in other households. First, the original recording system was too expensive and complex with cameras and microphones recording every room of the house - requiring a total of 11 cameras,

14 microphones and many terabytes of storage. Second, the recording system required that modifications be made to the home, with holes cut into ceilings for installing the cameras and microphones, and over 3,000 feet of wiring. Third, the recording system was not portable as it was integrated into the house. Finally, since the recording system was installed in the house of the principal investigator, it was relatively easy to manage privacy and there was little need for remote control and upkeep of the system since engineers were free to come and go into the house to fix and replace components. The SR is our attempt at a portable version of the HSP audio/video recording technology designed to address these limitations and be suitable for remote deployment, while still allowing for naturalistic and longitudinal recordings of the same quality. As reviewed in greater detail in this paper, the SR enables swift, cost-effective, and easy-to-use deployment in multiple home settings, and is organized in the following way. The first section describes our motivation and deployment scenario for the SR. The second section reviews the SR design. The third section covers SR hardware architecture, followed by on-board software in section four. The fifth and sixth sections review privacy and data management features, respectively. The final section presents our conclusions of this work and promising future directions.

2. MOTIVATION & DEPLOYMENT

2.1 Motivation

Autism Spectrum Disorders (ASD) include autism, Asperger syndrome, and atypical autism (otherwise referred to as Pervasive Developmental Disorder Not-Otherwise-Specified)[3], and affects 1 in 88 children in the United States[4]. Criteria for diagnosing ASD rely on behavioral features since reliable biological markers have yet to be established.

Currently, ASD is not diagnosed until 3-5 years of age[10], despite a growing body of literature suggesting that behavioral abnormalities are evident in the first two years of life. While a few case studies have been carried out (e.g., [8]), parents' retrospective reports (e.g., [17]) and analyses of home videos of children later diagnosed with ASD (e.g., [2]) are the primary methods used to identify behavioral features of ASD before 24 months.

When asked about initial concerns regarding their child with ASD, at least 30-50% of parents recall abnormalities dating back to the first year, including poor eye contact[11, 12, 25], lack of response to the parents' voices or attempts to play and interact[11, 12, 17, 25], and extremes of temperament and behavior ranging from alarming passivity to marked irritability[11, 12]. Similarly, several studies of early home videos have revealed behaviors indicative of ASD in children later diagnosed compared to those of typically developing children[2, 5, 15, 18]. During the first year of life, children with ASD are distinguished by a failure to orient to name[18, 15], decreased orienting to faces[18, 15], reduced social interaction[2], absence of social smiling[2], lack of spontaneous imitation[15], lack of facial expression[2], lack of pointing/showing[18, 15], and abnormal muscle tone, posture, and movement patterns (e.g., inactive or disorganized)[2]. Taken together, these findings indicate that disruptions in social, communicative, affective, regulatory, and motor domains are evident early in autistic children's development.

Although parents' retrospective reports and home video analyses clearly point to early abnormalities in an autistic child's development, this body of research is potentially limited by a host of methodological problems (as reviewed in [28]). First, a parent's incidental observations regarding subtle social and communicative differences may be limited compared to systematic assessments by trained clinicians[23]. Second, parents' tendency to use compensatory strategies to elicit their child's best behaviors (with or without awareness) may affect their behavioral descriptions[5]. Retrospective parental reports may also suffer from distortions of recall, especially when parents are asked to remember behaviors that occurred many years earlier. For instance, having already received a diagnosis of ASD for their child, parents may over-report behaviors that are consistent with the diagnosis. Retrospective reports are also likely to include significant inaccuracies with respect to the description and perceived timing of early behavioral signs. Finally, environmental manipulations and systematic presses for specific behaviors cannot be controlled in retrospective studies.

Home video analysis has significant strengths over retrospective parental reports as it allows the observation of behaviors as they occur in familiar and natural settings, and enable objective rating of behavior by unbiased observers. However, this methodology also has potential limitations. The primary shortcoming is that parents typically record videotapes to preserve family memories rather than document their child's behavior over time. As a result, tapes from different families vary as a function of the length of time the child is visible, the activities that were recorded, and the quality of the recording. Moreover, if children do not behave as expected or desired, parents may re-record taped segments until they obtain a more favorable response. Observations from home videos also vary considerably between children and depend on the particular contexts selected for taping. Another potential problem relates to the sampling contexts of home videotapes in so much as they may not have provided sufficient opportunity for social communicative behaviors.

Innovative audio/video systems that can easily be deployed in home settings are needed to better capture, quantify, and communicate early behavioral manifestations of ASD in order to refine early detection methods, elucidate developmental trajectories and neurodevelopmental mechanisms associated with the disorder, and refer children on the spectrum to early intervention programs during critical periods of development.

2.2 Deployment

From January of 2011 to July of 2011, a total of four SRs were deployed in households in Massachusetts, Rhode Island, and Connecticut. Of the four households, three had typically developing children and one had a child diagnosed with ASD, all 2yrs of age. The purpose of the deployment was to supplement a longitudinal NIH-funded study of language development in children with ASD led by researchers at the University of Connecticut to assess language comprehension; investigate the relationships between ASD children's early language development and their later language/cognitive outcomes; and determine how more detailed measures of on-line efficiency in language comprehension might predict ASD children's individual variation. Use of the SRs in this research was undertaken to provide an ecologically-valid, densely sampled (2-3 hours per day), and

extremely efficiently analyzed audio-visual corpus of each child's speech and home environment. This project is in line with recent studies of young typically developing children's motor[1] and language[16] development, which reveal that such dense sampling vastly increases sensitivity to the occurrence and non-occurrence of words/motor behaviors, thus rendering more accurately the patterns of development involved in their use. The audio from three of the households has been fully transcribed and preliminary results were presented at the International Meeting for Autism Research (IMFAR) in May of 2012[7].

In future deployments we envision deploying one SR per household. Its location will vary depending on the aims of any given study. Our current focus has been on the study of child development. For this purpose an SR is typically placed in a child's playroom, where the greatest amount of social interaction with other people can be repeatedly observed.

3. DESIGN

Given our motivation and deployment scenario, requirements for the SR design necessitated an observation system that was easy to transport, setup, install, and use without the requirement for home modifications (i.e., reduce invasiveness). The system also had to be inconspicuous and unobtrusive, without looking like a piece of "lab equipment." It was also designed to contain all necessary hardware in a modern furniture-like form factor easily adjustable for many room configurations.

Design inspiration drew from modern furniture lighting including overhead and cantilevered overhead lamps. A central base unit was chosen to house all electronics, hard drives, audio processing equipment, horizontal view camera, and touch screen controller in a secure and easy-to-access custom shelving system. A brushed stainless steel vertical boom extended from the base unit, curving slightly to meet a stainless steel curved horizontal boom that housed a vertical view camera near the middle of a room ceiling. Because of the large overhanging vertical camera boom, the system had a tendency to be unstable. As a solution, the vertical boom was designed to adjust through a manual crank system whereby one can brace the SR "head" against the ceiling, with additional adjustability for angled ceilings. The SR can be moved easily and installed in approximately 30 minutes, and once in place is very stable.

Initial designs for the SR included automated leaf-style lens cover systems to shield camera lenses during times when not recording, and a curved fiberglass base enclosure. However, to simplify the unit and cut costs, we replaced this design with a painted sheet metal cover and coverless camera lens system with an LED light indicating whether the device was recording or not. Over the course of the project, nine units were manufactured in Brooklyn, NY with custom designed parts fabricated at factories in Florida, Oregon, and New York.

4. HARDWARE ARCHITECTURE

The SR has a dual camera system, one overhead facing down and the other frontal facing horizontally. The cameras used in the recorders are "Lumenera Le165c" that transmits data over Ethernet. Both cameras are outfitted with "Fujinon FE185C057HA-1" lenses that have a 185-degree angle-of-view. This allows the overhead camera to capture most of the room from the top. The cameras are configured to record at 15 frames per second at a resolution of 960 by 960 pixels. Figure 2 shows sample downsized frames taken from the overhead and frontal cameras from one of the SRs. The resolution of the video is sufficient to identify human participants and surrounding objects. The audio sensors used in the recorders are "AKG C562CM" boundary layer microphones that use the surface in which they are embedded as a pickup. This allows a microphone placed in the head of the recorder to pick up speech in any corner of the room.

Figure 1: *The Speechome Recorder. Photo by Rony Kubat.*

To further reduce cost and invasiveness, and increase usability, the SR was designed to run for months without maintenance or technical support. The recorders were outfitted with a voltage regulator, cooling fans and UPS battery backup (allowing the recorder to run for up to 30 minutes without power). Additionally, we included an IP-Addressable power supply that allowed us to remotely turn power to any of the hardware components in the recorders on or off. Moreover, the recorders have on-board computers for data compression, and a touch display controller. Finally, the recorders have 4TB of disk storage, sufficient to hold about 60 days of continuous recordings (however, as noted in the section below, we can record continuously for much longer than 60 consecutive days using periodic remote data uploads).

5. ON-BOARD SOFTWARE

The SR employs both on-board and server-side software. For video/audio encoding we used tools developed for the HSP[9]. Other than the encoding tools, the on-board software can be divided into two categories: *Remote Upkeep* and

Figure 2: *Sample frames from the overhead and frontal cameras. Faces have been blurred in order to maintain privacy.*

User Interface.

5.1 Remote Upkeep

To enable the SR to run for months without the need for on-site visits, we wrote a comprehensive diagnostic software suite that runs on the recorders. This software automatically checks the operational status of all hardware components in a recorder and sends hourly reports to our off-site server. This, along with the ability to remotely control power with the IP-Addressable power supply, helps address possible issues at their outset (e.g., if the temperature inside a recorder is rising we can remotely shut the system down to cool it). The UPS battery backup allows the recorder to run for 30 minutes in the event of a power outage. During that 30 minutes, all data is backed up, a report is sent to our off-site server, and the recorder prepares for a clean shutdown by closing all applications and turning of non-essential hardware components. When the power comes back on, the recorder automatically starts itself and becomes fully operational in a matter of minutes.

To eliminate the need for swapping hard drives, data is transmitted over Internet to our servers during periods of inactivity (i.e., when the recorder is not being used). In this way, new data can be recorded over old data, ensuring that the 4TB disk storage in the recorders never get full, thus removing the need for manual drive replacement. Additionally, since data is transmitted daily to our server, researchers with appropriate permissions can access newly acquired data with only a few days delay.

5.2 User Interface

Figure 3 shows the user interface on the SR. The user interface on the touch screen controller was designed to be very simple, containing the following:

- *On/Off*: The top red button, marked with a camera icon, is the video-audio recording *"on/off"* button. It allows the user to turn recording on or off (the device itself is always on). The button turns green when recording and red when not recording.

- *Consent*: The left most button at the bottom, marked with a check mark, is the *"consent"* button. This button uses the front facing camera to take a picture of the person who is consenting to be recorded. The picture is sent to our off-site server and then used by an auditing officer to verify that only people who have consented have been recorded. This data auditing process is explained in detail in the *Privacy Management* section below.

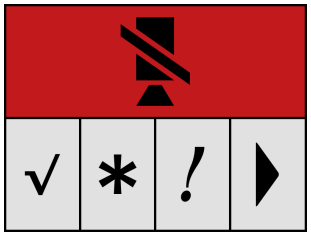

Figure 3: *User Interface of the Speechome Recorder.*

- *Ooh*: The second button from the left, marked with a star, is the *"ooh"* button. This button allows the user to event mark (i.e., create a time stamp in the video and audio record) recently recorded material, thus enabling researchers and/or the family to easily access interesting footage/data.

- *Oops*: The third button from the left, marked with an exclamation mark, is the *"oops"* button. This button allows the user to mark any recently recorded segment for deletion(e.g., in case something embarrassing was mistakenly recorded). Any segment marked in this way is automatically deleted without being seen by anyone else.

- *Playback*: The last button, marked with a play icon, is the *"playback"* button. This button allows the user to playback and review any recorded data using the recorder's screen. Figure 4 shows the playback interface. The calendar on the right highlights days when data was recorded. Once a day is selected, the user can playback clips from that day using the movie player on the left. The user can also switch between overhead and frontal cameras. If the user wishes to share any of the clips with others they grant permission, they can select the "Share this clip" button to mark the clip. As we will describe in the *Data Management* section, users have access to a website that allows them to view and download their clips.

6. PRIVACY MANAGEMENT

Privacy management was a very important factor in our design of the SR. With the HSP, the principal investigator and his family was the focus of recording, this is not true of SRs. The "oops" button on the recorders described earlier was borrowed from the HSP to allow for local control of the data by the users. The button opens a dialog box that allows users to specify any number of minutes of recorded data and retroactively and permanently delete it from disk (both local and server-side).

More importantly however, we needed to make sure not to record anyone who did not consent to be recorded. When a recorder was installed, parents signed consent forms for themselves and their children. Moreover, consent was also obtained from frequently visiting grandparents, therapists, and friends. Using the "consent" button on the recorders,

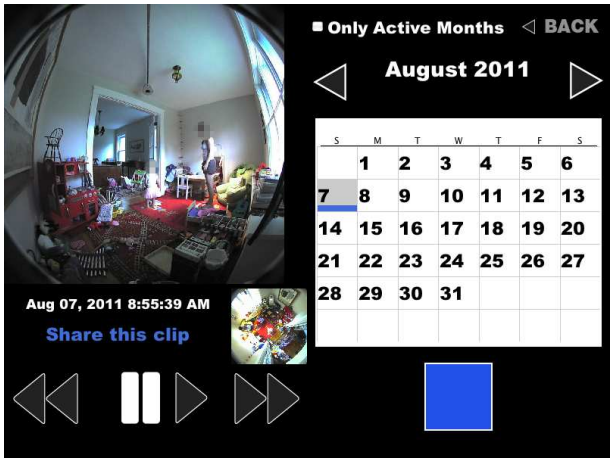

Figure 4: *The playback interface of the Speechome Recorder. Faces have been blurred in order to maintain privacy.*

pictures were taken of all consented participants. In order to make sure no unconsented individuals were recorded, an independent *data-auditor* was hired to review captured data on a weekly basis. The job of the data-auditor was to go over all data recorded that week for each of the households and check people in the video against pictures of consented individuals in that household.

However, due to the large amount of data recorded every week, it was impractical to have the data-auditor manually inspect all videos in full. To overcome this problem we developed a semi-automated auditing system that allowed the data-auditor to efficiently review collected data.

Figure 5 shows the processing pipeline of our semi-automated auditing system. The system works by first running the video data through an automatic face detection algorithm, using frontal camera video. This face detection system uses OpenCV's[6] automatic face detection algorithm and is trained on sample faces from different angles (front, profile, etc). The algorithm detects frames in which there is at least one human face. Our system then randomly picks one of these frames for every 15 seconds of recorded video, with the assumption that the recording of any one person would be at least 15 seconds in length. These frames, along with the consent pictures, were then passed to our auditing software to be used by the data-auditor.

Figure 6 shows the auditing software. The numbered buttons on the top allow the data-auditor to select between SRs in different households. Once a recorder is selected, frames detected by our automated system are shown (the left image in Figure 6) alongside a list of pictures of consented people from that household (the image scroll-panel on the right side in Figure 6). The data-auditor can then compare the faces in the frames to the consent pictures. If a face is not found in the consent pictures, the auditor can use the giant red "X" button to mark that picture. When completed, a report is sent to the researchers about all marked images. If a non-consented person was identified, we could either try to get the person's consent or, if that fails, delete all recorded data where that person is present. These audits were performed once a week on all data. Over our six months of deployment

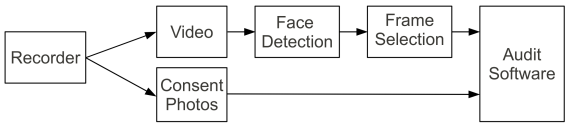

Figure 5: *Data processing pipeline for the auditing software.*

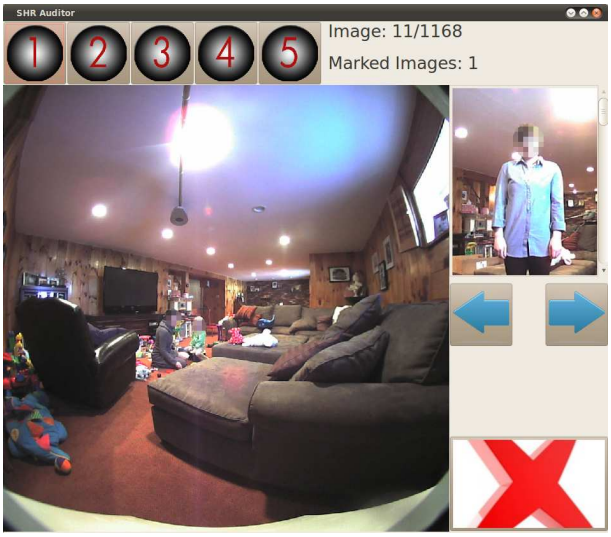

Figure 6: *Data auditing software for the Speechome Recorder. Faces have been blurred in order to maintain privacy.*

recording, auditing revealed no instances of errors in which recordings were made of unconsented individuals.

7. DATA MANAGEMENT

A SR generates a large database of multimedia content. In this section we describe how we manage and process this massive dataset. As mentioned earlier, when the recorders are inactive (i.e., not being used) data is transmitted over Internet to our servers. The data transmitted is lossless compressed video and raw audio, both time stamped and time-aligned. Once the data reaches our servers it is immediately backed up on multiple drives. After that, the raw data goes through various pipelines described below.

7.1 Data Processing

Before any data processing is undertaken it goes through the auditing process described earlier. After privacy auditing, the audio is processed and transcribed. Figure 7 shows the complete audio processing pipeline. The audio is first passed through a noise reduction filter using the "Audacity" speech processing toolkit[24]. Next, the audio is passed to an automated speech detection system (which uses a boosted decision tree classifier[20]) that categorizes the audio into speech and non-speech segments. The speech segments are then passed through an automated speaker identification system[20] (trained on the dataset) to identify the speakers in each speech segment. The speech segments are also passed to a state-of-the-art, semi-automatic transcription

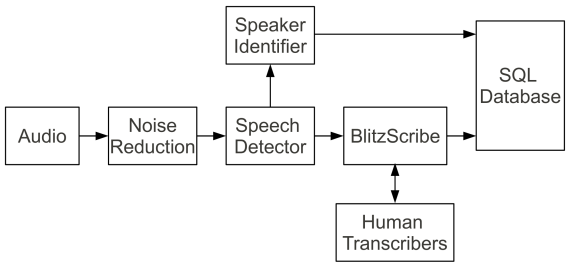

Figure 7: *Audio processing pipeline.*

tool called *BlitzScribe*[1][20] prior to manual transcription. The transcriptions along with the speaker identifiers and time-stamps are saved in a central SQL database.

Next, the raw data from the three channels (audio, front-view video, top-view video) are time aligned and converted to a standard mpeg movie format (keeping the raw data). The movie is then time-aligned with the transcriptions. Since data collected by an SR are recorded using the same technology as the HSP, many of the visualization tools developed for that project, such as *TotalRecall*[2][13] can also be used for SR recorded data.

7.2 Data Access

Three primary groups of people access data captured by a SR: engineers who maintain the system; families who have recorders installed in their homes; and scientists who analyze collected data. In order to facilitate data access to these parties we created a secure, password protected website that serves as a front-end user tool. Figure 8 shows a simplified overview of the data pipeline from the recorders all the way to the end users: the families, engineers, and scientists. Below we describe what type of data each of these groups accessed.

7.2.1 Families

One of the motivations for families to participate in this project was to have access to all data they recorded with the SR. In order to facilitate that, all the raw audio and video collected was converted to an mpeg movie format and uploaded to a movie player on our secure website for family access. Figure 9 shows a screenshot of the movie player. The player comes with an interface that allows families to lookup videos based on days and times they were recorded. All the recordings, including clips selected for sharing, could be viewed using this movie player. The families also had the option of saving any of the video segments to their local drives.

7.2.2 Engineers

The engineers involved in this project needed to have access to data to maintain the recorders and servers. The video player described above allowed the engineers to review the quality of audio and video feeds (e.g., to see if lenses are focused) and make sure they are time-aligned. Moreover, the engineers received nightly reports from the recorders on their operational state (as described earlier in the *Remote Upkeep* section). These reports included information on different operational aspects of the recorders, such as the amount of data recorded that day, the average temperature of the recorders, the status of different hardware parts in the recorders, and more.

[1]BlitzScribe is a tool for manual speech transcription that is up to six times faster than other available transcription methods[20].

[2]TotalRecall is an audio-video browser and annotation system. It provides a global view of the corpus indicating when and where recordings were made. The system also supports limited types of speech transcription and video annotation.

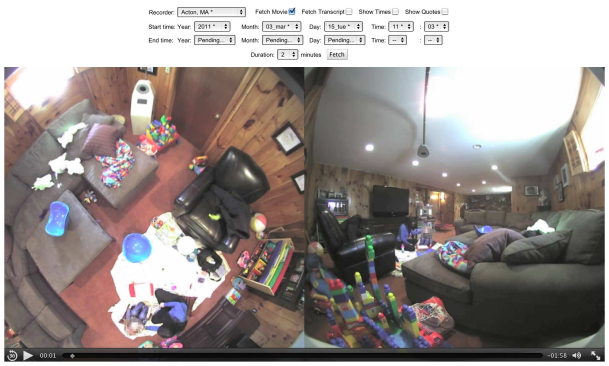

Figure 9: *The Speechome Recorder movie player. The player shows the video feed from the top and front cameras. Users can search for and play videos from specific dates and times. Faces have been blurred in order to maintain privacy.*

7.2.3 Scientists

As mentioned previously, the main goal of the SR project was to record longitudinal and naturalistic data to study child language development and developmental disorders. Other than being able to watch the recordings using the movie player, researchers analyzing the data could read the transcripts associated with the video they were viewing (Figure 10). The transcripts were displayed in a format compatible with CLAN[14], a standard analysis tool used by child language researchers. Moreover, the researchers could automatically divide the recordings into continuous segments of recorded data showing one or more activities (e.g. play time, therapy sessions, etc). In addition to being able to search the recordings by date and time, the researchers could also search for clips containing particular words or phrases (e.g. all instances of the word "water").

Additionally, researchers had access to weekly transcription reports to assess transcript progress and generate various language statistics. Moreover, the researchers had access to usage data reports (Figure 11) that allowed them to monitor the use of the recorders and, if needed, to encourage families to use the recorders more or differently, and to address any questions. It is also worth noting that due to our mostly automated processing pipeline, it took on average two days for any recorded data to go through the pipeline and for all information to be available on the website (excluding transcriptions) for viewing.

8. CONCLUSIONS

The main contributions of the SR are addressing privacy management issues and reducing cost, complexity, and invasiveness of naturalistic longitudinal recordings to facilitate swift, cost-effective deployment in home settings. With

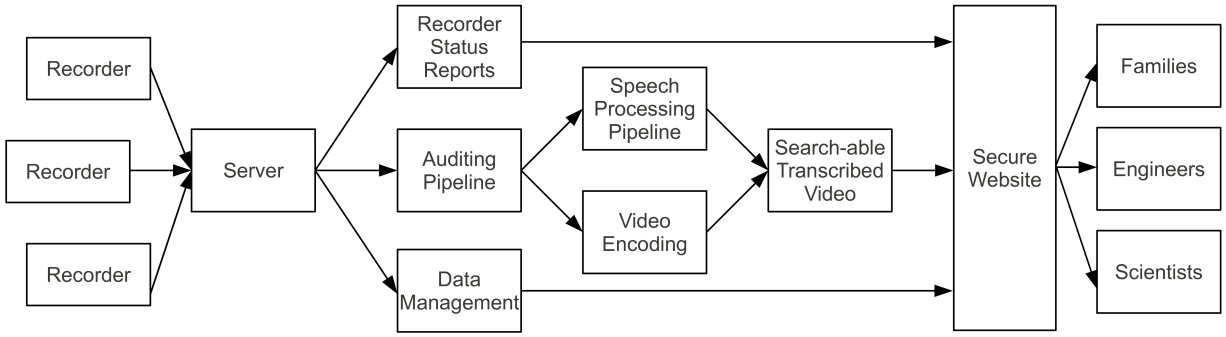

Figure 8: *A simplified overview of the data pipeline from the "field"(households) all the way to the end users: families, engineers and scientists.*

Transcript for Acton, MA, 2011-04-05 10:22:00 - 2011-04-05 10:24:00.

Speakers:
CHI: Chi
FT: Ft
TWI: Twin

@CHI: fell on your leg
@CHI: this is me
@FT: oh, you're the teacher this time?
@CHI: yeah
@FT: okay
@CHI: i be the jj
@CHI: do this
@CHI: jj this
@FT: is this the same?
@CHI: no!
@FT: is this the same?
@CHI: no!
@FT: is
@FT: was trying to jj
@FT: this one the same?
@CHI: yeah!
@FT: i did it. yay. now, my turn to be the teacher
@CHI: i want to be

Figure 10: *Sample transcribed session. Some information has been modified in order to maintain privacy.*

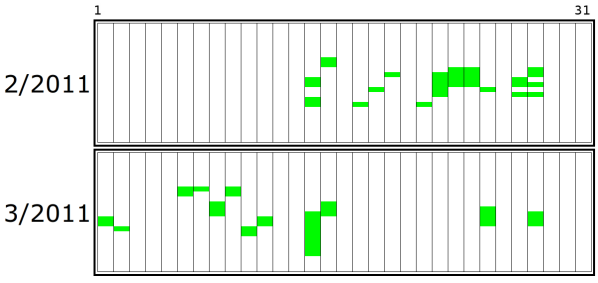

Figure 11: *Sample data report from one recorder. Each row represents a month and each column represent a day in the month. Green blocks mark the times when the Speechome Recorder was being used.*

the help of the SR, ultra-dense, longitudinal studies of child development can now be scaled to include multiple households and participants, allowing for generalization of findings between multiple children and developmental comparisons across clinical and non-clinical populations. Moreover, the recording, processing, and visualization pipelines enabled by this technology allow for data to be transferred from the "field" (i.e., households) and processed and analyzed by researchers with unprecedented speed.

With rapid advancements in video capture, processing, and storage capabilities currently underway, the SR can be further enhanced to reduce size, cost, and complexity. Also, in the future, the SR could be modified to receive additional data streams from wireless, wearable physiological sensors that may shed light on internal biological processes involved in speech and behavioral development observed in audio and video channels. Finally, future work could explore the use of the SR to facilitate real-time tele-consultation applications, bridging families in the home with allied health professionals at a distance.

On a final note, building the SR commenced a few years ago. The focus of the project was to develop engineering, privacy, and data access management methods for making ultra-dense, longitudinal recordings in multiple home settings. However, with rapid advancements in technology, any such recording technology will not be practical from an engineering point of view for more than a few years. Future recording technologies will likely be as small as a light bulb, require little to no installation and cost very little. However, the scientific practicality of this project is independent of its engineering practicality. The scientific approach employed in this project opens a new door of data analytics on a scale not seen before, and is already producing interesting research results using similarly obtained datasets[27, 19, 26].

9. ACKNOWLEDGMENTS

We would like to thank all the families who participated in this work. We also thank Joe Wood for help installing the recorders; Philip DeCamp, Rony Kubat, Matt Miller, Brandon Roy, and George Shaw for help with various technical aspects of the project; and Prof. Letitia Naigles and Devin Rubin for help recruiting families to be recorded.

This work was generously funded by grants from NIH (R01 2DC007428) and the Nancy Lurie Marks Family Foundation.

10. REFERENCES

[1] K. E. Adolph. *Encyclopedia of Infant and Early Childhood Development*, chapter Motor/Physical Development: Locomotion. San Diego, CA: Academic Press., 2008.

[2] J. L. Adrien, C. Bathelemy, A. Perrot, S. Roux, P. Lenoir, L. Hameury, and D. Sauvage. Validity and reliability of the infant behavioral summarized evaluation (IBSE): A rating scale for the assessment of young children with autism and developmental disorders. *Journal of Autism and Developmental Disorders*, 22:375–395, 1992.

[3] A. P. Association. *Diagnostic and statistical manual of mental disorders (4th ed.)*. Washington, D.C: APA, 1994.

[4] J. Baio. Prevalence of autism spectrum disorders: autism and developmental disabilities monitoring network, 14 sites, United States, 2008. *Surveillance Summaries*, 61 (SS03):1–19, 2012.

[5] G. Baranek. Autism during infancy: A retrospective video analysis of sensory-motor and social behaviors at 9-12 months of age. *Journal of Autism and Developmental Disorders*, 29:213–224, 1999.

[6] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[7] I. Chin, D. Rubin, A. Tovar, S. Vosoughi, M. Cheng, E. Potrzeba, M. Goodwin, D. Roy, and L. Naigle. Dense recordings of naturalistic interactions reveal both typical and atypical speech in one child with asd. In *Proceedings of the International Meeting for Autism Research*, Toronto, Canada, (2012, in press).

[8] G. Dawson, J. Osterling, A. N. Meltzoff, and P. Kuhl. Case study of the development of an infant with autism from birth to two years of age. *Journal of Applied Developmental Psychology*, 21:299–313, 2000.

[9] P. DeCamp. Headlock: Wide-range head pose estimation for low resolution video. MIT M.Sc. thesis, 2007.

[10] P. Filipek, P. Accardo, G. Baranek, E. Cook, G. Dawson, B. Gordon, J. Gravel, C. Johnson, R. Kallen, S. Levy, N. Minshew, B. Prizant, I. Rapin, S. Rogers, W. Stone, S. Teplin, R. Tuchman, and F. Volkmar. The screening and diagnosis of autistic spectrum disorders. *Journal of Autism and Developmental Disorders*, 29:439–484, 1999.

[11] C. Gillberg, S. Ehlers, H. Schaumann, G. Jakobsson, S. O. Dahlgren, R. Lindblom, A. Bagenholm, T. Tjuus, and E. Blinder. Autism under age 3 years: A clinical study of 28 cases referred for autistic symptoms in infancy. *Journal of Child Psychology and Psychiatry*, 31:921–934, 1990.

[12] Y. Hoshino, M. Kaneko, Y. Yashima, H. Kumashiro, F. R. Volkmar, and D. J. Cohen. Clinical features of autistic children with setback course in their infancy. *Japanese Journal of Psychiatry & Neurology*, 41:237–245, 1987.

[13] R. Kubat, P. DeCamp, B. Roy, and D. Roy. Totalrecall: Visualization and semi-automatic annotation of very large audio-visual corpora. In *Ninth International Conference on Multimodal Interfaces*, 2007.

[14] B. MacWhinney. The childes project: Tools for analyzing talk. Lawrence Erlbaum Associates, Mahwah, NJ, 3rd edition, 2000.

[15] A. E. Mars, J. E. Mauk, and P. W. Dowrick. Symptoms of pervasive developmental disorders as observed in prediagnostic home videos of infants and toddlers. *Journal of Pediatrics*, 132:500–504, 1998.

[16] L. R. Naigles, E. Hoff, and D. Vear. Flexibility in early verb use: Evidence from a multiple-n diary study. *Monographs for the Society for Research in Child Development*, 74, 2009.

[17] M. Ohta, Y. Nagai, H. Hara, and M. Sasaki. Parental perception of behavioral symptoms in Japanese autistic children. *Journal of Autism and Developmental Disorders*, 17:549–563, 1987.

[18] J. Osterling and G. Dawson. Early recognition of children with autism: A study of the first birthday home videotapes. *Journal of Autism and Developmental Disorders*, 24:247–257, 1994.

[19] B. C. Roy, M. C. Frank, and D. Roy. Relating activity contexts to early word learning in dense longitudinal data. In *Proceedings of the 34th Annual Cognitive Science Conference*, 2012.

[20] B. C. Roy and D. Roy. Fast transcription of unstructured audio recordings. In *Proceedings of Interspeech*, Brighton, England, 2009.

[21] D. Roy. New Horizons in the Study of Child Language Acquisition. In *Proceedings of Interspeech 2009*, Brighton, England, 2009.

[22] D. Roy, R. Patel, P. DeCamp, R. Kubat, M. Fleischman, B. Roy, N. Mavridis, S. Tellex, A. Salata, J. Guinness, M. Levit, and P. Gorniak. The Human Speechome Project. In *Proceedings of the 28th Annual Cognitive Science Conference*, pages 2059–2064, Mahwah, NJ, 2006. Lawrence Earlbaum.

[23] W. L. Stone, E. L. Hoffman, S. E. Lewis, and O. Y. Ousley. Early recognition of autism: Parental reports vs. *clinical observation. Archives of Pediatrics and Adolescent Medicine*, 148:174–179, 1994.

[24] A. Team. Audacity (version 1.3.4-beta) [computer program]. http://audacity.sourceforge.net/, 2008.

[25] F. R. Volkmar, D. M. Stier, and D. J. Cohen. Age of recognition of pervasive developmental disorder. *American Journal of Psychiatry*, 142:1450–1452, 1985.

[26] S. Vosoughi. Interactions of caregiver speech and early word learning in the speechome corpus: Computational explorations. MIT M.Sc. thesis, 2010.

[27] S. Vosoughi, B. C. Roy, M. C. Frank, and D. Roy. Contributions of Prosodic and Distributional Features of Caregivers' Speech in Early Word Learning. In *Proceedings of the 32nd Annual Cognitive Science Conference*, 2010.

[28] L. Zwaigenbaum, A. Thurm, W. Stone, G. Baranek, S. Bryson, J. Iverson, A. Kau, A. Klin, C. Lord., R. Landa, S. Rogers, and M. Sigman. Studying the emergence of autism spectrum disorders in high-risk infants: Methodological and practical issues. *Journal of Autism and Developmental Disorders*, 37:466–480, 2007.