BabyCloud, a Technological Platform for Parents and Researchers

Xuân-Nga Cao, Cyrille Dakhlia, Patricia Del Carmen,
Mohamed-Amine Jaouani, Malik Ould-Arbi, Emmanuel Dupoux
Laboratoire de Sciences Cognitives et Psycholinguistique,
EHESS / Ecole Normale Supérieure / PSL Research University / CNRS / INRIA
Paris, France
ngafrance@gmail.com, dakhliacyrille@gmail.com, pat_delcarmen@hotmail.com,
jaouani.mohamed.amine@gmail.com, m.ouldarbi@gmail.com, emmanuel.dupoux@gmail.com

Abstract

In this paper, we present *BabyCloud*, a platform for capturing, storing and analyzing day-long audio recordings and photographs of children's linguistic environments, for the purpose of studying infant's cognitive and linguistic development and interactions with the environment. The proposed platform connects two communities of users: families and academics, with strong innovation potential for each type of users. For families, the platform offers a novel functionality: the ability for parents to follow the development of their child on a daily basis through language and cognitive metrics (growth curves in number of words, verbal complexity, social skills, etc). For academic research, the platform provides a novel means for studying language and cognitive development at an unprecedented scale and level of detail. They will submit algorithms to the secure server which will only output anonymized aggregate statistics. Ultimately, *BabyCloud* aims at creating an ecosystem of third parties (public and private research labs...) gravitating around developmental data, entirely controlled by the party whose data originate from, i.e. families.

Keywords: language acquisition, cognitive development, linguistic development, day-long recordings, child speech corpus, child activity corpus

1. Introduction

During their first years, infants develop both physically and cognitively at an amazing speed: many parents are eager to know more about this general process and about how well their child is doing. Recent progress in infant speech database collection (Xu et al., 2008; Roy, 2009; VanDam et al., 2016; Casillas et al., 2017; Warlaumont et al., 2017) and computational modeling of developmental processes (Martin et al., 2016; Ludusan et al., 2015; Carbajal et al., 2016a; Carbajal et al., 2016b; Ludusan et al., 2017) open up possibilities to measure and model the progress of children using data collected in their natural environment. In other words, it becomes technically possible to offer parents scientifically grounded analytics and tools to explore and document their child's progress.

In this paper, we introduce *BabyCloud*, a platform for capturing, storing and analyzing day-long audio and photo recordings of children's linguistic environments. The platform is structured around 4 components: (a) *Baby Logger* which collects high definition pictures and high-quality recordings to be later decoded by automatic speech recognizers, (b) *Baby Dock* which transfers the data to a secure cloud server, recharges the recorder's batteries, (c) *Baby SmartBox*, a secure cloud database which stores and indexes raw data and extracts metadata and developmental analytics (child's linguistic landmarks, linguistic and social interactions clips and analyses) using machine learning algorithms and (d) *Baby Explorer*, which consists itself in two sub-components: a mobile application which allows families to manage and search their data, and an API accessible to scientists, will let them query summary statistics of the data and remotely run algorithms (no access to raw data).

2. Related Work

One of the more comprehensive resource for child language acquisition research is the CHILDES database (MacWhinney, 2000), which, though a tremendous source of data for language studies, presents some shortcomings. The corpora were primarily constructed to document the child's productions, and only secondarily their input; as a result, many recordings are done in the laboratory setting, and are not entirely representative of his or her activities and input in their natural environment. In addition, the majority of the corpora contain only orthographic transcriptions, many of the audio recording were done for human transcription and the quality is not ideal for automatic speech processing.

An ambitious project for gathering day-long home audio recordings was initiated at MIT with the Human Speechome Project (Vosoughi and Roy, 2012). It consisted in a continuous recording of all of the child's environment (audio and video), some of which has been transcribed. However, due to the private nature of the data, it is not accessible to outside researchers.

HOMEBANK (VanDam et al., 2016) created by the DARCLE[1] group, is another repository for day-long family audio recordings and addresses some of the limitations encountered with the CHILDES database or the Human Speechome Project. It gathers large datasets of audio (and sometimes video) recordings of the infant in his or her natural environment. It addresses the privacy problem by a legally binding protocol of nondisclosure agreement signed by the researchers, their institutions and the repository. This only partially addresses the problem however, given

[1] Day-long Audio Recording of Children's Linguistic Environments

once the agreement is signed, the data can be copied onto the researcher's computer, with no possibility to control what becomes of the data. This is a problem as data protection laws differ across countries (for example, the General Data Protection Regulation law will be enforced in May 2018, aiming at protecting and securing all European residents' personal data, while the US Data Privacy regulation ensures only online consumer's data security and privacy is under FTC authority since 2000).

Another technical issue for the automatic processing of large quantities of speech data is the audio quality of the initial recordings. One recorder developed with this objective in mind is the LENA[2](Xu et al., 2008; Oller, D.K, 2011). The device has been miniaturized and engineered to maximize the quality of the recording, while worn by children of various ages. The LENA system also provides an analysis package using machine learning technology to automatically segment the recordings in audio chunks labelled with broad classes (target child, adult male near, background noise, etc). One main limitation of this tool is that the data and processing tools are proprietary and only optimized for English households, making it difficult to adapt to other languages or projects. In addition, it offers no search tools for the parents.

3. The *BabyCloud* Platform

We developed *BabyCloud*, an innovative open-source platform that strives to connect two communities of users: families and academic researchers. It stands out from existing models with its commitment to protect, first and foremost, the collected data and to give the full control and ownership to the party whose data originate from: the families.

Many parents who contribute to science by donating their infant's data and their time also care about their child's development and his/her prospects at school. To address this, the platform will enable the development of novel services such as the ability to follow the development of their child on a regular basis through language and cognitive analytics (growth curves in number of words, verbal complexity, social skills, etc.) even before he/she attends school. In addition, they will have complete control of their data usage, with the option to open up null/partial/full data access to researchers and/or third parties.

For academic research, the platform provides a novel instrument for studying language and cognitive development, potentially, at an unprecedented scale and detail level. Unlike existing data repositories, our platform ensures simultaneously total privacy protection and openness: the data are encrypted and never leave the secure server; however, researchers are able to submit algorithms to the server to analyze the data and only anonymized aggregate statistics will be allowed to get out of the server (see figure 1 for the *BabyCloud* workflow).

[2]https://www.lena.org/

3.1. The *Baby Logger* Component

The *Baby Logger* is a light, ergonomic and wearable recorder. It has been designed to combine efficiency and convenience for daily life usage. The device records sounds thanks to an array of eight high-quality microphones, captures high resolution pictures and data from a three-axis accelerometer. It is protected by an open source hardware licence, and can be reconfigured for particular projects (see figure 2 for more details).

3.1.1. Operating Modes

The operating modes of the device optimize ease of use and respect of privacy. Two versions of the device have been designed, one for the parents and the other for the child. They have the same features, but the one worn by the parents stops recording and goes into hibernation mode when they go far away from the child. This proximity detection is possible thanks to a radio-frequency identification (RFID) chip affixed on the parents' recorder whereas the child's one only possesses a tag. This setup gives more privacy to the parents and protects the child from a permanent wave exposure. In addition to this automatic setting, a manual privacy mode can be selected, with the parents temporarily stopping the recording on both devices. This situation may occur when the conversation contain sensitive data or other guests are present and don't want to be recorded. The camera will also be equipped with a flap to hide the lens if the family doesn't want the *Baby Logger* to take pictures. The decision to add the camera is to have pictures from the child's perspective, which can shed light on his/her home environment and development.

The user interface is very simple with only two buttons, one for turning on/off the device and the other, easily accessible for selecting the privacy mode. Two LEDs indicate each one of the following status: privacy vs recording mode, and a warning for docking station operation (low battery or full memory). The *Baby Logger* is equipped with a large internal memory to allow a maximum of 24 hours of continuous sound recordings and a stream of images every 10 seconds. Less dense acquisition settings are available, notably, random or periodic 2 to 10 minutes sample. The battery has been designed to last for at least 24 hours under the maximum recording density.

3.1.2. Choice of Technology

The choice of the electronic components has been made to optimize data quality and power usage. Our recorder is built around a high performance microcontroller, equipped with a Cortex-M7 ARM microprocessor providing enhanced performance and optimized power consumption, representing the most relevant solution for IoT and multimedia systems. Audio acquisition is performed with an array of 8 MEMS (Micro ElectroMechanical Systems) microphones, a new generation of digital microphones directly built on semiconductors and integrating digital circuits (ASIC). MEMS are an attractive solution because of their small size and weight, allowing systems to be miniaturized. The camera is a 5 Mega Pixels CMOS digital camera. It has a Digital Signal Processor (DSP) and

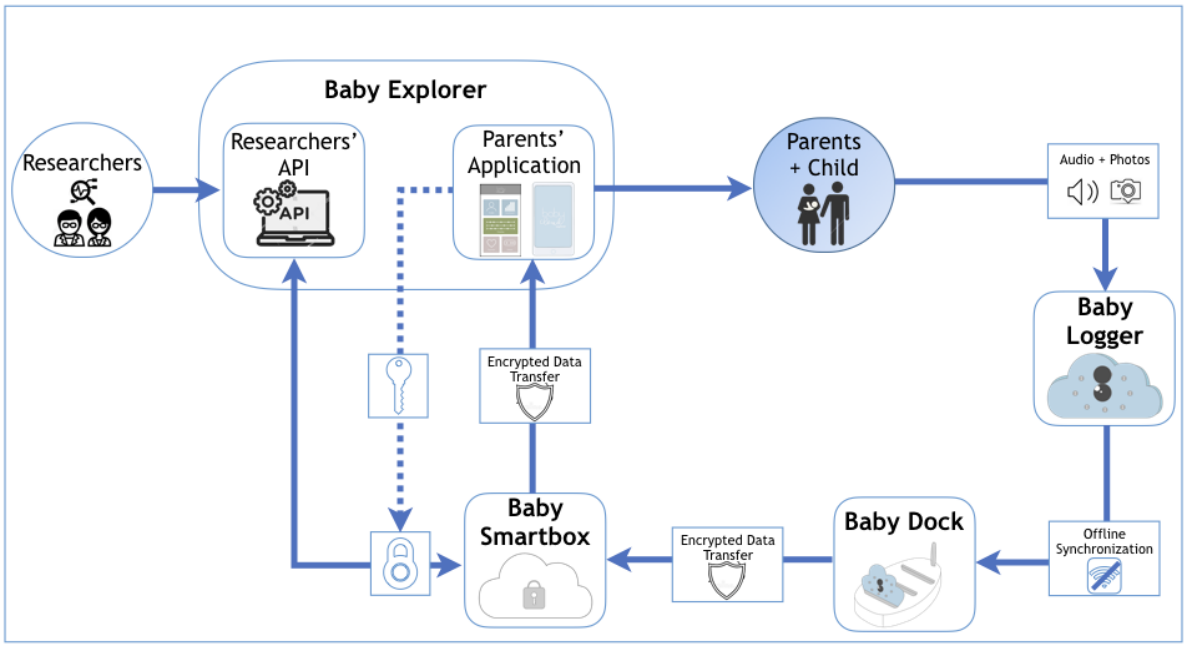

Figure 1: *BabyCloud* Workflow.

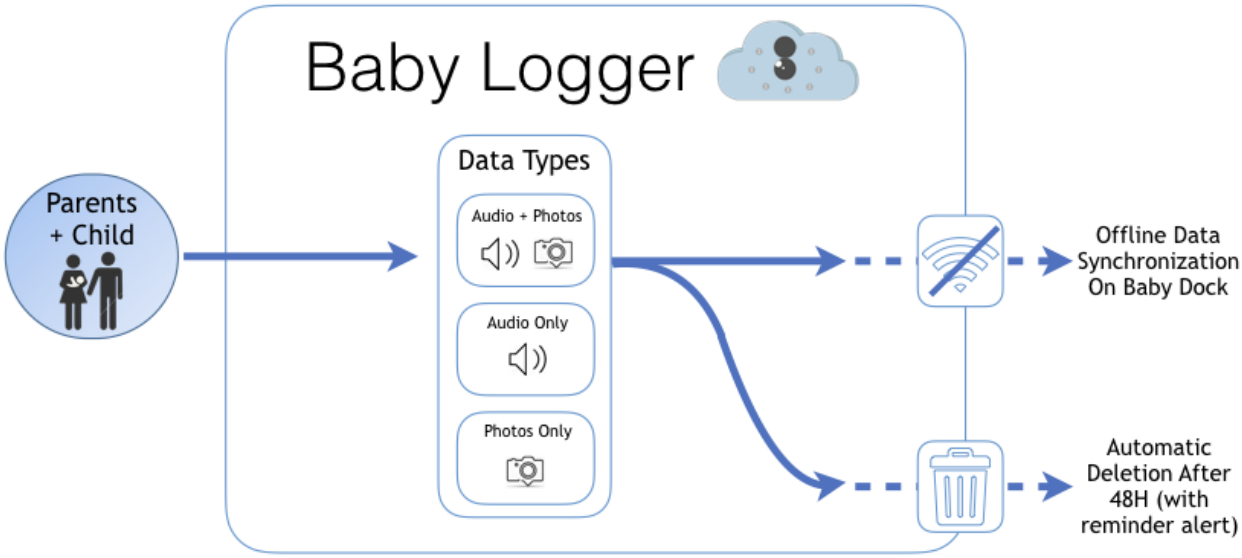

Figure 2: *Baby Logger* Architecture.

presents a control interface allowing customization of the camera settings.

3.2. The *Baby Dock* Component

While the device is recording, no data is being transferred to save on battery life and to protect the child from wave exposure. The wireless data synchronization is performed only when the battery is being recharged onto the *Baby Dock* (see figure 3 for more details).

The *Baby Dock* represents the interface between the wearable recorder and the secure cloud. A Raspberry Pi 3 (our preferred solution but any computer with a Wifi module working as a host for a local network can be used) is connected to the Internet thanks to an ethernet wire which ensures the link with the home network. A three-phase cyclical software will run on the *Baby Dock*:

- Recorder synchronization Phase: During this stage, the connection to the Internet is disabled to ensure the security of the raw data being transferred from the recorders to the *Baby Dock*. The transfer is done instead via the local Wifi network setup between the Life Logger and the Dock. Simultaneously the Life Logger's batteries are recharged through inductive charging. At the end of synchronization, the data is removed

from the *Baby Logger*'s memory.

- Data Filtering Plugin: Signal processing algorithms are applied offline to the data in order to separate speech from background noise (source separation, speech enhancement) and segment the recordings according to two classes (speech, non-speech). This significantly reduces the amount of data transferred to the cloud. We will use open-source sytems (Jin and Schultz, 2004; Rudzicz, 2013) which will be tuned to the baby logger data and customized to fit into the baby dock.
- Cloud synchronization Phase: The last phase of the cycle ensures the transfer of the encrypted data to the secure cloud server via an ethernet connection and their removal from the Dock.

3.3. The *Baby Smartbox* Component

The *Baby Smartbox* stores the collected raw data, runs advanced signal processing and machine learning algorithms to automatically add meta-data (speaker diarization, activity detection, estimation of the child's vocal maturity, etc), and provides a database server. The server is protected to provide access only to authorized users via a login authentification process. It interfaces with the Baby Explorer via a RESTful API developed within the Python Flask framework (see figure 4 for more details).

This component will have a core secure database management system selectively granting access to data to different classes of users (parents, researchers). In addition, it will have a set of machine learning algorithms (plugins) which will automatically data annotations at various linguistic levels. We are currently focusing on replicating some of the annotation layers of the LENA system, using state-of-the-art, open source and retrainable software. These include:

- Segmentation broad class speaker ID. This requires algorithms that perform Speech or Voice Activity Detection: they segments speech from background noise (Uhle and Bäckström, 2017). Broad class speaker ID classifies the resulting segments into a small number of categories, like target child, other child, males, females.
- Speaker diarization. This task is more difficult than the above, because it requires to classify utterances according to an unknown number of speakers. The task is made easier if the speakers can be 'enrolled' in advance, i.e. that a few minutes are manually annotated. The current performance of speaker diarization systems based on I-vectors is dependant on utterance length. Typically, performance is not very good with short utterances (Kanagasundaram et al., 2016).

Other plugins will be added as they become available in open source format. In particular, we are working in collaboration with the ACLEW[3] consortium. This is funded by a transatlantic research initiative (Digging Into Data project[4]) aiming at creating a common annotation scheme for diverse (culturally and linguistically) infant datasets, and at building tools to semi-automatically analyze those datasets. The plug-ins currently under development are:

- Syllabic segmentation. Syllables are acoustically saillant events in speech and can be efficiently detected (Räsänen et al., 2018). In turn these can be used as proxy for utterance length.
- Vocal maturity. Infant produce several stages of vocalization at different developmental stages (vegetative vocalizations, crying, laughter, canonical babbling, varigated babbling, etc.). Given their rather stable shape across languages, it seems feasible, using standard machine learning tools (markov models, SVM or random forest models) to construct a retunable classifier that will enable to produce developmental statistics based on these categories.
- Child-directed vs. adult-directed speech. Parents in many culture address their infants in specific ways (Fernald and Kuhl, 1987), and child directed input seems to be a better predicted for language acquisition than total input (Weisleder and Fernald, 2013). Being able to automatically annotate this difference in register could help predictive models of language development.

As these plugins are developed and open sourced, they will be incorporated into the platform. We construct all of these modules to be trainable. Which means that the automatic annotation can be retuned to specific recordings, provided the parents gave authorization for a subset of the data to be annotated manually. In that case, the annotated data is split into a training set and a test set to evaluate the reliability of the machine annotation. Similarly, manual annotations are required to adapt any new plugin to the particular data collected by the baby logger. All of the automatic annotations are considered derived data and added to the raw data. They are used to derived analytics, such as daily vocal activity, linguistic complexity (mean utterance length), and social responsiveness indexes (turn taking), which will be accessible to the parent for their child and to the scientist as aggregate statistics (see section 3.4.2).

3.4. The *Baby Explorer* Component

The *Baby Explorer* is designed to cater to 2 types of users: parents and researchers.

3.4.1. The *Baby Explorer*: parents' interface

It is a hybrid mobile/web application developed within the Ionic framework using the following technologies: Typescript (Javascript) and Angular 4 for data processing and actions linked to screen/hardware events, HTML5 and CSS3 for graphic user interfaces, Cordova for hardware access to use mobile native functionalities, and PostgreSQL for databases.

[3] Analyzing Child Language Experience around the World: https://sites.google.com/view/aclewdid/home

[4] https://diggingintodata.org/

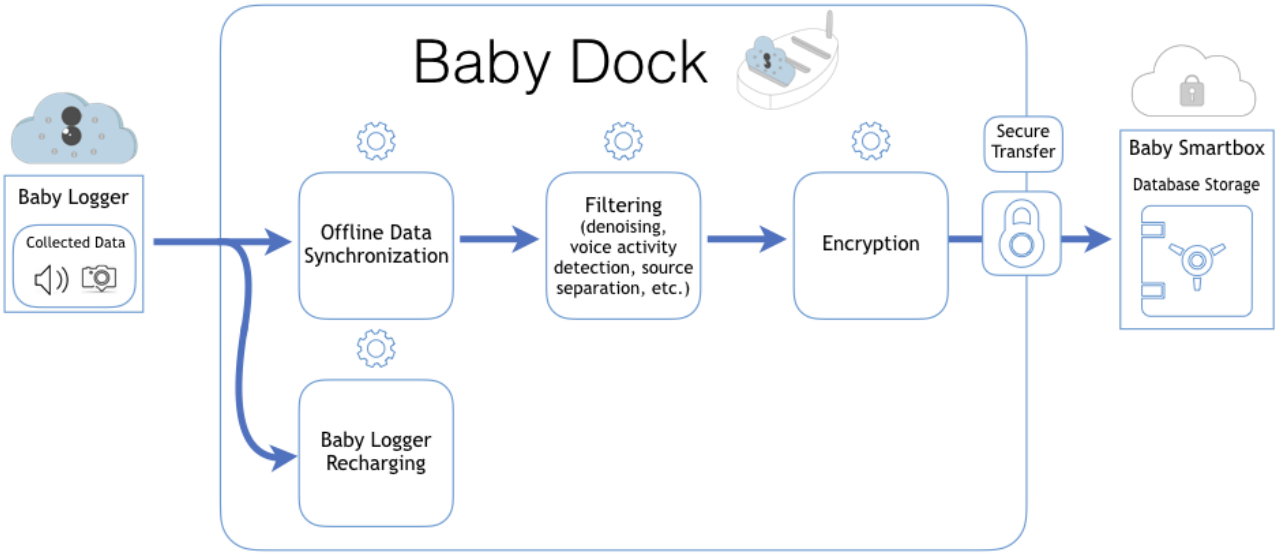

Figure 3: *Baby Dock* Architecture.

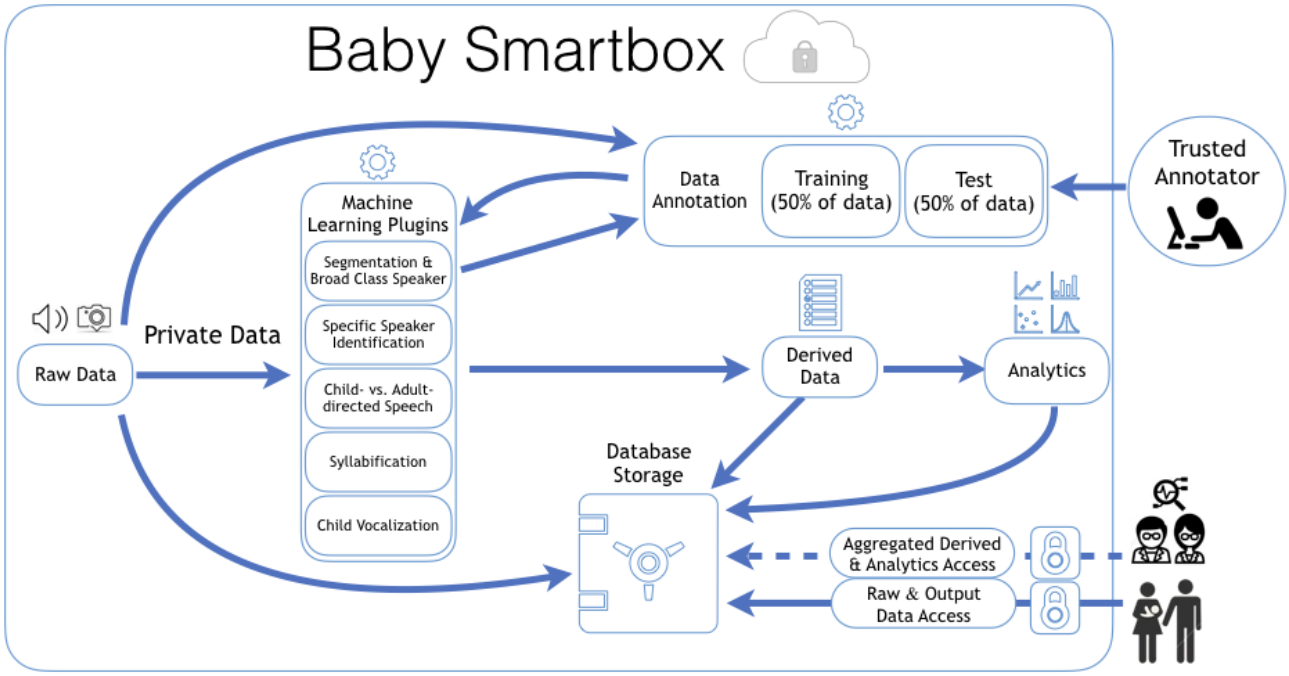

Figure 4: *Baby Smartbox* Architecture with example machine learning plugins.

The objective is to actively engage families with their dataset by offering them an attractive tool with several functionalities: (a) browse their child's recordings using time and/or activity filters, (b) track their child's language and cognitive development with the provided analytics and statistics, (c) manage the data access for researchers (see figure 5 for more details).

To ensure that parents fully understand the extent of their rights with regards to data protection, an online consent form must be signed when parents log in for the first time onto the *Baby Explorer*. Unlike traditional terms of agreement clauses, this one is aimed at protecting the user's rights and privacy and making sure that the user understands the terms of contract he/she is agreeing on. This last step is verified by an online quiz while the user is going through the consent form. Since parents own the data, they have the option to update, erase, remove them or ask for their portability from the secured cloud, using the interface. In the same way, the child, once he/she turns 13 years old (based on Children's Online Privacy Protection Rule aka "COPPA"), will be required to express his/her free consent

to continue or stop his/her participation to the project.

- Data View: The data view interface helps families browse through their infant's data via a customized and ergonomic time-line, allowing to simultaneously access audio recordings and their associated pictures. On this timeline parents can control the audio and image channels (play, stop, rewind, re-scale, etc), and can filter the data by activity. They can bookmark their favorite recordings or pictures, share them with relatives or friends and set privacy settings on their data with regards to their use for research. The pictures taken from the child's device will inform parents and researchers on his/her perception of his local environment (people he/she is looking at, location, activities such as playing, eating, napping...).
- Dashboard: The dashboard displays an overview of the metrics about the child's linguistic development. It allows parents to follow his/her cognitive development on a daily basis through analytics and statistics (growth curves in number of words, verbal complexity, social skills, etc.).
- Access Authorization: This interface allows families to see all the requests made by researchers to access their data. Parents are able to read information about the enquirers such as their name, contact information, status, summaries of their projects and collaborations, publications, etc. Acceptance from the parents is a pre-requisite for scientists to work on the data. Parents have the option to define presets to automatically accept/reject/block certain types of requests (educational research, medical research, etc.). They can also grant/revoke access on a project by project basis.

3.4.2. The *Baby Explorer*: researchers' interface

This API (Application Programming Interface) offers a bundle of functionalities and tools dedicated to researchers. They will have at their disposal new and large dataset with fine layers of annotations, ready for analysis and obeying the regulations on data protection. Beforehand, all researchers will be required to pre-register their studies and get approved by an ethics committee. Once their algorithms have been tested with public data on the virtual machine sandbox at their disposal, scientists can send access requests to parents to run these algorithms on their meta-data. When the access is granted, they have to use the validated virtual machine, designed to make sure that the raw data cannot be retrieved. The only output retrieved by the researchers is anonymized statistics of the data. In a later part of the project, we will add the possibility for scientists to request algorithmic access to the raw data in order to add new layers of meta-data (see figure 6 for more details).

4. Completed Work

At the time of writing, we have completed the following steps: we have a working prototype of the Baby Logger, of the Baby Dock, of the Secure Cloud Database and of the Baby Explorer parent's interface. The algorithms for speech analysis are being developed in parallel by the ACLEW team (see section 3.3 above). We have submitted the project description to our local ethics committee. It has been reviewed favorably and will be approved after revisions. The proposal has also been submitted for a formal authorization to the National Committee for data protection. It is currently under review. The miniaturization and baby friendly design of the Life Logger is underway, and the scientist's interface under study.

We also conducted a short pilot with six families to apprehend the parents' needs and expectations. These families were selected from our babylab's database which comprises more than 1000 volunteer families. The infant wore a T-shirt with a pocket hiding the audio recorder (such as the LENA or a USB key recorder). The recording happened at home for one/two day(s). We asked for the parents' feedback afterwards about the experiment. All were in favor of pursuing the study but they expressed reservations on the recording duration (once or twice a week instead of continuous recordings), were interested in a more practical and fit-all recorder. There were mixed reactions on the camera function (some parents wanted to disable it completely). We annotated (with the parents' consent) a small sample of the collected data in order to integrate the analytics into the application. A second interview with the families will be scheduled to show the results on the application and have their feedback on the user experience to improve on the functionalities of the platform.

Once the protocols are validated by the data protection and ethics committees, we plan to scale up the project in terms of number of families (20 families) and duration (one year or more). We aim to achieve a functional platform with agile software/hardware development through rapid iteration cycles. During these cycles of improvement, we will pay particular attention to user experience and data protection.

5. Expected Impact and Conclusion

In this paper, we presented an innovative platform aimed at reinforcing collaboration between parents and researchers. The benefits of such a platform for the community users and for society are diverse.

The benefits for families are twofold. The first one is health related: the language and cognitive analytics could help parents spot potential developmental delays and trigger early medical intervention. The platform does not propose medically validated diagnostic tools but parents may use the provided data and consult speech and language therapists to get further assistance and medical advice. A secondary benefit is related to the quality of life. Our platform will provide cloud storage space and indexing services related to the child: automatically extracted pictures and audio snippets. In addition, parents will be able to include other types of information (potentially,

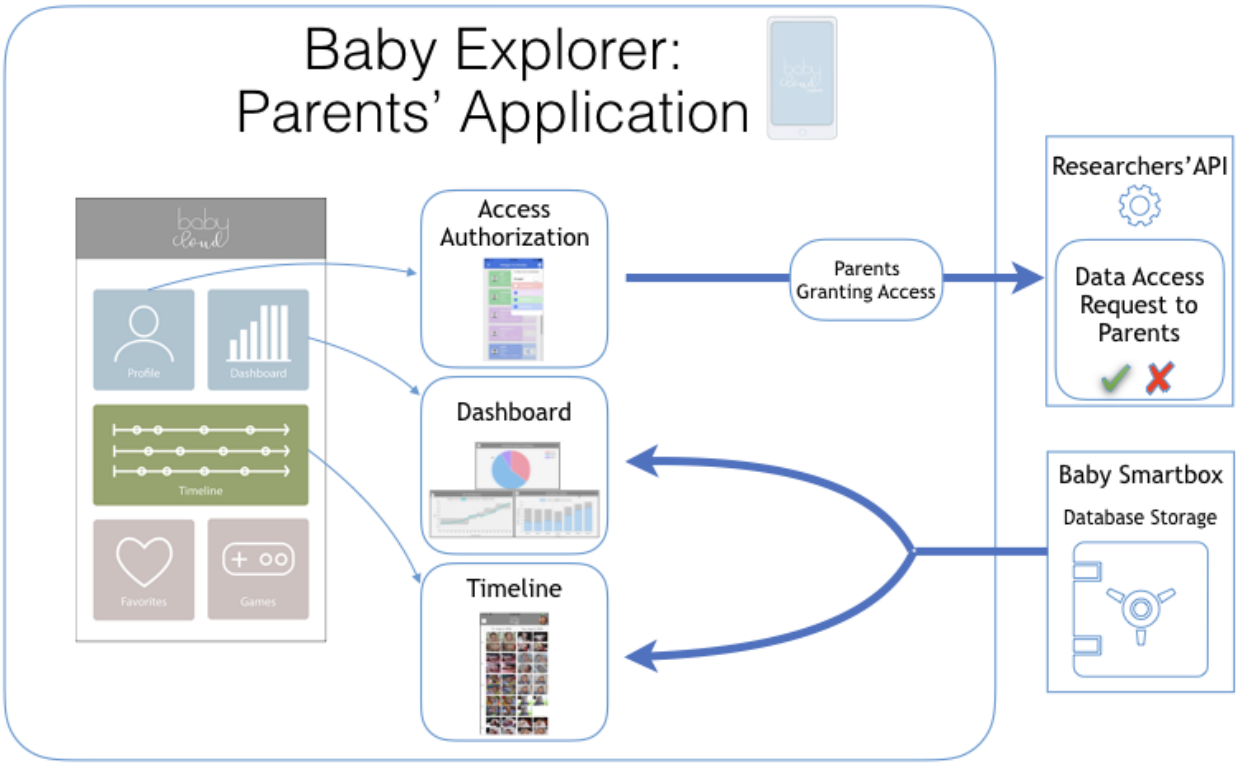

Figure 5: *Baby Explorer* Architecture: Parent's Application.

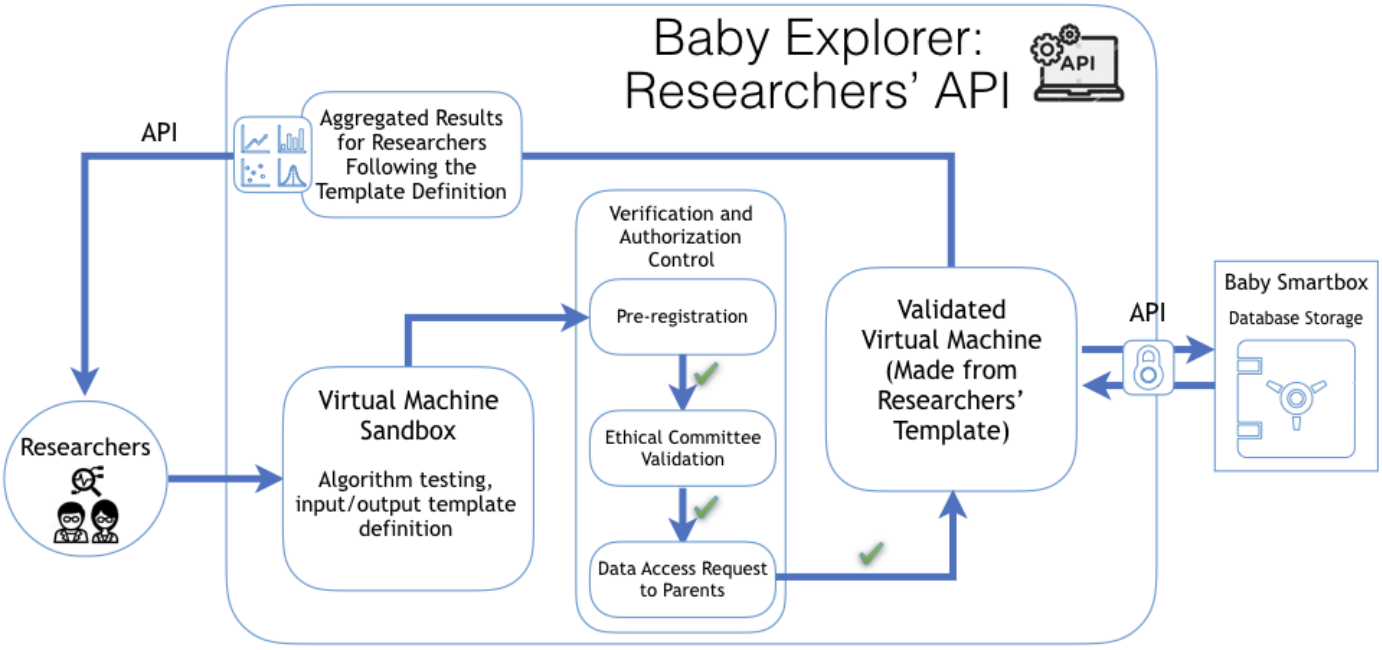

Figure 6: *Baby Explorer* Architecture: Researcher's API.

school and medical records), which security level will be guaranteed. The whole portfolio will constitute a digital asset for the child, which can be permanently accessible as a memory of early childhood.

For academia, large recordings of infant development will allow researchers to establish new quantitative and predictive models at a level that was not possible before.

These models will feed back into better analytics and in a long term, detect early signs of potential cognitive delays. An additional benefit will be the involvement of parents into science (citizen science), who will help collect data and perhaps, volunteer annotations.

In a later phase, we plan to include a third community to interact with the families and researchers: private companies. For private IT companies, personal data is a valuable source for big data analytics and building business strategies based on the customer's preferences. The platform will offer access to large and rich amount of data but which will be protective of families' privacy. Under this strict privacy preserving protocol, companies would be able to offer families new tools and analytics. Vice-versa, families may open up, through smart contracts, partial aspects of their data to companies for specific purposes. These smart contracts will remain under the lab's supervision to ensure that the family's privacy and data protection are safeguarded.

Finally, society may benefit from the platform in the area of developmental pathologies. Developmental disorders are typically diagnosed too late (i.e., at a time when the child is already lagging behind in school), resulting in mental strains and heavy financial costs for families and also for society. Tools that will raise awareness about these disorders and their impact may help alleviate those costs.

6. Bibliographical References

Carbajal, J., Dawud, A., Thiollière, R., and Dupoux, E. (2016a). The 'language filter' hypothesis: Modeling language separation in infants using i-vectors. In *EPIROB 2016*, pages 195–201.

Carbajal, J., Fér, R., and Dupoux, E. (2016b). Modeling language discrimination in infants using i-vector representations. In *The 38th Annual Conference of the Cognitive Science Society*, pages 889–896.

Casillas, M., Bergelson, E., Warlaumont, A. S., Cristia, A., Soderstrom, M., VanDam, M., and Sloetjes, H. (2017). A new workflow for semi-automatized annotations: Tests with long-form naturalistic recordings of childrens language environments. In *Proc. Interspeech 2017*, pages 2098–2102.

Fernald, A. and Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant behavior and development*, 10(3):279–293.

Jin, Q. and Schultz, T. (2004). Speaker segmentation and clustering in meetings. In *8th International Conference on Spoken Language Processing*, page 597–600.

Kanagasundaram, A., Dean, D., Sridharan, S., and Fookes, C. (2016). Dnn based speaker recognition on short utterances. *arXiv preprint arXiv:1610.03190*.

Ludusan, B., Cristia, A., Martin, A., Mazuka, R., and Dupoux, E. (2015). Learnability of prosodic boundaries: Is infant-directed speech easier? *Journal of the Acoustical Society of America*, 26(3):341–347.

Ludusan, B., Mazuka, R., Bernard, M., Cristia, A., and Dupoux, E. (2017). The role of prosody and speech register in word segmentation: A computational modelling perspective. In *ACL 2017*.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates, 3rd edition.

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., and Cristia, A. (2016). Mothers speak less clearly to infants: A comprehensive test of the hyperarticulation hypothesis. *Psychological Science*, 140(2):1239–1250.

Oller, D.K. (2011). Lena: automated analysis algorithms and segmentation detail: how to interpret and not overinterpret the lena labelings. In *LENA Users Conference*, Denver, CO.

Räsänen, O., Doyle, G., and Frank, M. C. (2018). Prelinguistic segmentation of speech into syllable-like units. *Cognition*, 171:130–150.

Roy, D. (2009). New horizons in the study of child language acquisition. In *Proceedings of Interspeech*.

Rudzicz, F. (2013). Adjusting dysarthric speech signals to be more intelligible. *Computer Speech Language*, 27:1163–1177, 09.

Uhle, C. and Bäckström, T. (2017). Voice activity detection. In *Speech Coding*, pages 185–203. Springer.

VanDam, M., Warlaumont, A. S., Bergelson, E., Cristia, A., Soderstrom, M., De Palma, P., and MacWhinney, B. (2016). Homebank, an online repository of daylong child-centered audio recordings. *Seminars in Speech and Language*, 37:128–142.

Vosoughi, S. and Roy, D. (2012). An automatic child-directed speech detector for the study of child language development. In *Interspeech 2012*, Portland, Oregon.

Warlaumont, A. S., VanDam, M., Bergelson, E., and Cristia, A. (2017). Homebank: A repository for long-form real-world audio recordings of children. In *Proc. Interspeech 2017*, pages 815–816.

Weisleder, A. and Fernald, A. (2013). Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary. *Psychological Science*, 24(11):2143–2152, November.

Xu, D., Yapanel, U., Gray, S., and Baer, C. (2008). The lena language environment analysis system: the interpretive time segments (its) file. Technical Report No. LTR-04-2, LENA Foundation Technical Report.

7. Acknowledgements

Our research was funded by the European Research Council (ERC-2011-AdG 295810 BOOTPHON). It was also supported by the Agence Nationale pour la Recherche (ANR-10-IDEX-0001-02 PSL and ANR-10-LABX-0087 IEC) and the ENS Fondation (chaire Almerys).