Report for summer school CPS2018

LINGYUAN YANG
linyan@kth.se

July 31, 2018

Executive summary

This report is for the cyber security and privacy summer school 2018 in Trento. The name of our project is EspioNo. We provide device to secure your private conversation from listening by smart speakers. With a state-of-art technology disturbs the microphone but not human ear. To introduce our project, this report is divided into 4 sections. The first section will be focusing on the process of idea generation and optimization, as well as the technical basement. The second section will be introducing our business canvas, detailize our business structure. The third section will introduce our time line and expected cash flow. And the last section will evaluate the performance of myself both as a participant of summer school and a member of a team.

Contents

1 Problem and solution

1.1 Problem addressment

From the point that Amazon post its 90 seconds 'Alexa Loses Her Voice' advertisment, people begin to fascinate with smart speakers for the convenience it introduce into our life. However, to interacte with the users, smart speaker has to keep listening to them. This feature arise doubt and panic on privacy issue. Some reports publised recently seems to confirm that such concern make sense. According to The New York Times [1] and Quartz [2], alexa sometimes do record undesired conversation and even send it to other people. If this happens, why can't they tap us and make money by our privacy? Such tapping problem is a big worry for people who care about privacy. And the patent of Google and Amazon It's also been heard that when people talk about something they like or they want, an relative advertisment will be forwarded to them soon. A study has been done across America showing that 54% of the adults of the United States do not own a smart device in their home. 6.4 milion people said specifically they would not buy a smart speaker due to privacy reasons. 6.4 million people. And what about the 20 million people who want to increase their privacy. In the united states alone, we have not even talked about the rest of the world. And this problem will only increase in the future.
In this case, the problem we are going to solve is that people want to keep privacy while using smart speaker without undermine the convenience.

1.2 Problem validation

We use two ways to validate our problem. The first one is talking with people. The second way is reading reports and statistics about such problem. Firstly, we talk to people who has a smart speaker, 4 of our relatives are involved into our interview and all of them confirm that they do have such concern and they are willing to pay for products solving such trouble. We also happen to have "interivews" with Mr Jovan senior and junior during technical consulation and montoring slot. They both think Amazon and Google are not trustworthy and would like to have solutions for such problem, they represent the group of people who show interest about smart speaker but don't buy them due to privacy concerns.
We also collect statistics to validate our problem. According to a research conducted by Pew Research Center, 46% of the American adults have ever use digital voice assistants while 8% say they have ever use a stand-alone device such as Amazon Echo or Google Home. In the mean while, according to a survey from Business Insider [3] 40% of people are concerned about connected-home devices tracking their usage. And according to Business Wire , 6.4 million people said specifically they would not buy a smart speaker due to privacy reasons.

1.3 Solution and validation

1.4 Proposed solutions

To solve the addressed problem, we brainstorm and come up with several solutions and validate them one by one. Generally speaking, there are three ways preventing smart speaker from tapping.

- **Cut off power supply when not used** This way is to make a smart switch with speech recognition function to shut down the smart speaker when you don't want to use it anymore.

- **Monitor/Control the internet connection when not used** This way is to stand in between remote server and local speaker monitor all the internt traffic to secure no sensitive data been transmitted.

- **Physically prevent smart assistants from listening** This way is to disturb or deafen the microphone on smart speaker, make them unable to heard from us. And remove the interference when desired.

1.5 Validation process

In the beginning, these three solution all looks hopeful. But two of them are deprecated after we reading technical literature and doing test.

- **Why we abandon power switch solution** This solution is easy to validate. There is no doubt that it will work. But the problem is how well it works and if it's convenient. We do have a google home to conduct experiment. After we turn it off and reboot, it takes about 1 minute to be responsive again. Which means if we are going to use such solution, every time our customer want to use smart assistant they will have to wait for 1 minute after they say the key word.
  Customer buy smart speaker for the convenience. If our product makes life harder, it does't make sense and nobody will pay for it. In this case, we kill this solution first, even before look into technical detail.

- **Internet monitor solution** This solution actually include two different solution, internet switch and package sniffing. The second one is the inproved version of first one.
  The internet switch is what we originally come up with during brain storm. It works just like power switch, but instead of control over power, it controls upload traffic, prevent package being transmitted when not desired. The best thing about this solution is that it eliminates the drawback of booting latency. However, as long as we don't know the mechanism behind, we cannot ensure that the tapped audio won't be transmitted once reconnected to the internet.
  With such problem, we consult Yvo Desmedt during a consulation slot we booked. He suggest us to use sniffing technology to monitor what is actually transmitted. In other words, our product will act as a transfer center of package from smart speaker and analize the package to figure out if undesired package is being transmitted. With this suggestion, we do research and find out the audio data being transmitted are encrypted [4], which make sniffing work extremly hard. And it turns out somebody has already do simmilar job and we can derectly refer to it. According to a report [5] by Maik Morgensternfrom AV-TEST, the communication channel of Amazon Alexa uses TLS1.2 encryption with certificate validation/pinning. This technique prevented them to use a Man-In-The-Middle-proxy and read along the encrypted traffic. So they creat a scenario to evaluate the intensity of data transimission and estimate what is being transmitted. The scenario strats with 8 seconds of silence and then Asking Alexa "What time is it?" (6 seconds), then another 8 seconds of silence and followed by "Alexa – tea, earl grey, hot" and await the answer (5 seconds) finally the scenario end with 23 seconds of a normal conversation between two people (not including any Alexa keywords).

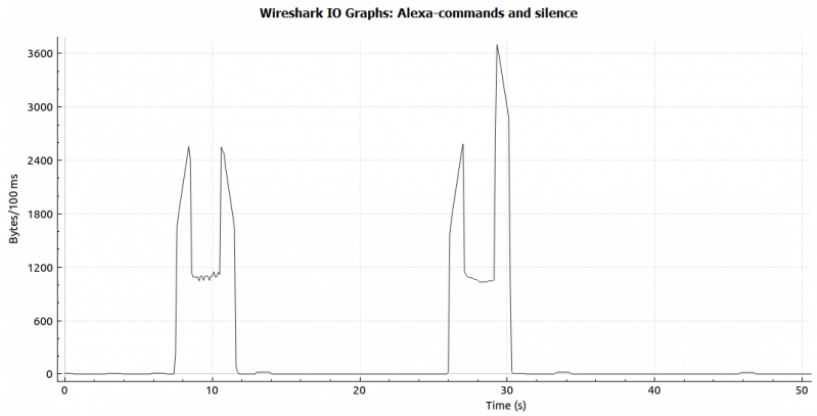

Figure 1: Transmitted bytes in the given scenario

We can see the result from figure 1 that the data transimission during conversation looks simmilar with silent time, it is very likely that Alexa do not send data if it's not working. This analisis is unfavorable to our solution.

Also, Prof. Yvo Desmedt give us an artical [6] about homomorphic encryption. With such technology we will be able to processing data, such as speech recognition, on remote server encryptedly wothout let the server konw a peice of information. To make use of such advanced technology we need to build our own smart speaker and the remote server to conduct homomorphic encrypted speech recognition. These is not viable for starting point, we plan to include it in our long term plan.

- **Audio disturbing solution** Interference is the first solution we come up with. The way how it works is to decrease the Signal Noise Ratio(SNR), which describe the level of a desired signal to the level of background noise, as low as possible. In other word, make noises to make conversation incomprehensive. There is one thing to be noted, a startup company called Privacy Shield has already
  - The first version of disturbing solution: use mini speakers to play white noise, which is a very common noise type, near the microphone of smart speaker. However, there is a big problem. The noise we play might bother our customer. According to the Acoustic Modeling report of google home [7], it could work well at SNRs ranging from 0 dB to 20 dB or above. So in the most optimistic scenario, we need to decrease the SNR to 0 dB which means the noise is as loud as human voice when reaching the microphone. This is load enough to be haerd several meters away.
  - The second version of our solution: use a case to cover the smart speaker which could both decrease the volume of human voice reaching microphone and the volume of noise needed. A case could fix the problem in the first version almost perfectly, it both decrease the volume of noise we need and the volem of noise leak out. But we need a mechanical structure to open and close the case, which should also be audio-controled. The way it work is shown in figure 2. This structure is viable although it require complicated mechanical structure to support smooth opening and closing of the case and also speech recognition function to remain the user experience. Not good enough but we still use this solution for our first pitch.

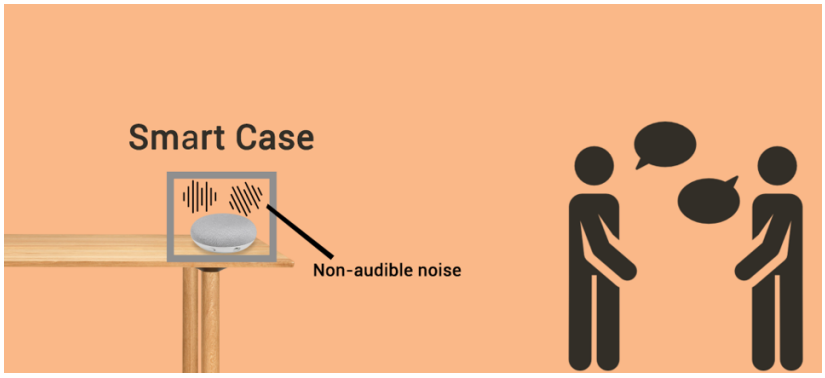

Figure 2: Audio disturbing solution

1.6 Final solution

The final solution actually evolved from the second version of Audio disturbing solution. After struggling on simplify our product and trying to add encryption elements to the solution. Fortunately, we found a newly published artical from UIUC [8] which provide a way to use inaudible ultrasonic sound to disturb microphones.
Inside the microphone, all the physical audio will first been transfered to electronic signal, after that the signal should pass a component called pre-amplifier which will amplify the analog signal to make sure it can be effectively measured by ADC. In this procedure, the output of pre-amplifier is linear only when input is in the audible frequency range which is 10-23 KHz, outside this range, the response exhibits

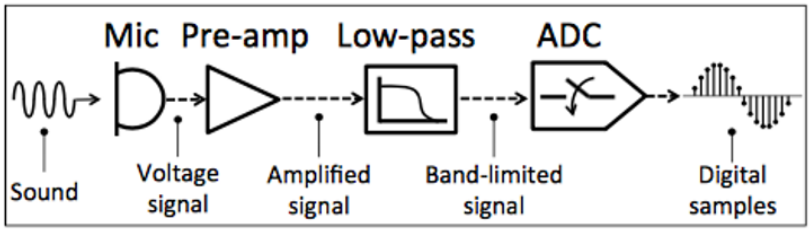

Figure 3: Circuit inside microphone

non-linearity which will introduce harmonic. In the artical, by making use of such non-linearity, they design the input ultrasonic sound and make a controlable shadow noise signal within the audible range, which has been perfectly recorded by microphone.

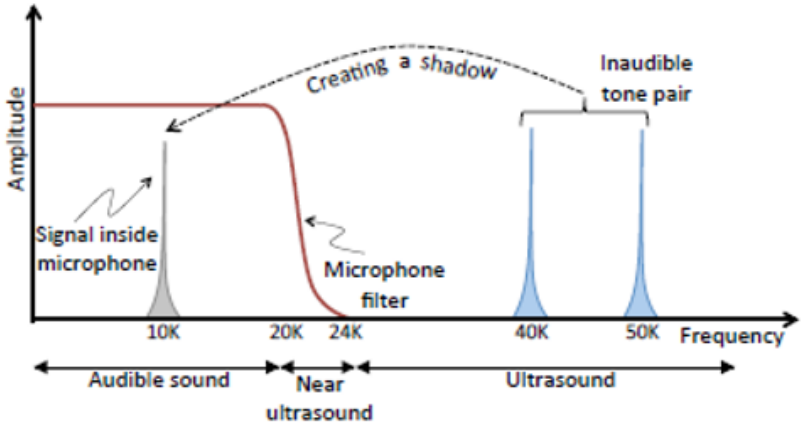

Figure 4: Use ultrasound to generate shadow noise

And by analyze the output signal from pre-amplifier, the researchers make sure that the ”shadow” noise is as expect and strong enough to disturb the speech recognition result. Being more specific, they use two tones at say 40kHz and 50kHz, When these tones arrive together at the microphone’s power ampli
er, they are ampli
ed as expected, but also multiplied due to fundamental non-linearities in the system. Multiplication of frequencies f1 and f2 result in frequency components at $(f_1 - f2)$ and $(f_1 + f_2)$, $(f_1 - f2)$ is 10kHz in this case. The harmonic measured is shown in figure 5.

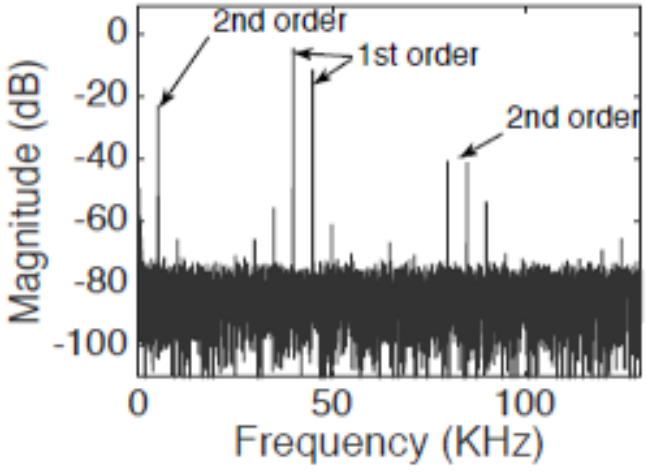

Figure 5: Non-linear Harmonic

This solution is perfect for the problem we addressed, however, we still facing questions during technical consulation.

- **Will the ultrasound hurt people?**
  The first obvious concern is about health and well-being, people are always worry about radiation and altrasound hurt their health. Actually ultrasound effect your body just as normal sound, frequency here does not make any difference. Only power of the voice matters [9]. According to the measurement conducted by the auther[8], we know that the amplifier have a very large magnitude, which is -40 dB/Hz around frequency of 45kHz, this means we can make our disturbing altralsound really weak to fulfill the desired effect.

- **Will the ultrasound hurt pets?**
  We know that the audible range of pets like dogs and cats are different from people, usually higher. What if the frequency of ultrasound is out of human audible range but fall in animal hearing range? This concern make sense. We have choice to design the frequencies of input ultrasound as long as their difference remian the same, however, the further the frequency deviates from 50kHz, which is a peak for magnitude, the smaller the amplification. This problem can be solved, although not perfectly right now.

- **This technology is brand new, is it business viable?**
  Our solution s based on a paper published last year. Such new technology could help us get rid of much competition, but also rise the problem that if it can be realize. According to the paper the authers already have a prototype for testing, and by reviewing the blog of the authers, we found out the prototype they made 6. It's a speaker array to disturb the microphone in a room, while the power of it is only 2W.
  We also write email to cantact with them. Their technology is pending for IP but they welcome usage for research and also open for business coorperation.

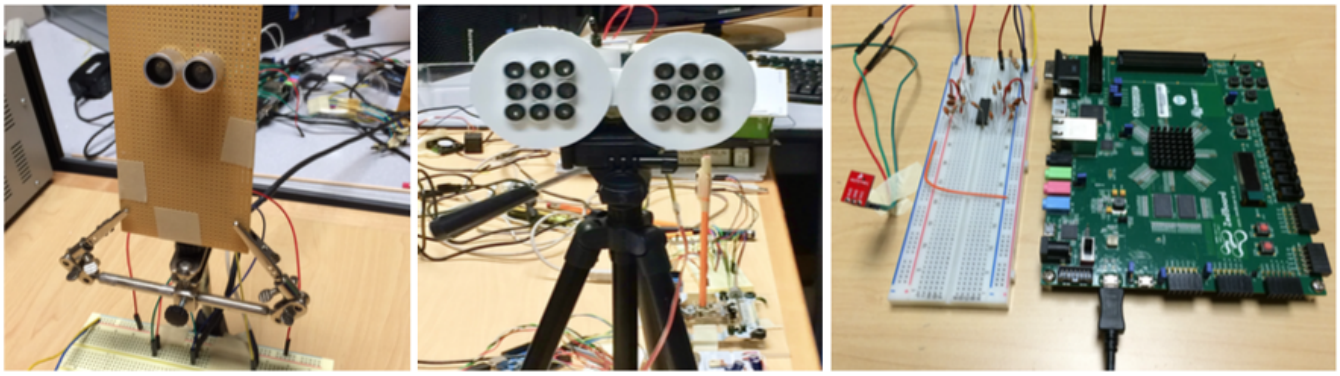

Figure 6: Prototype

After all of these problem we finalize the desining of our product. Which is shown in figure 7

2 Business modelling and planning

In our case, business modle is relative straightforward. We develope product and sell them. Which is called directly sale modle.

2.1 Business modelling

- **Key Partners**
  - **The authers of article** [8]. As long as the technology we rely on is based on others IP, or very likely have an IP. We need to coorperte with them. Not only for technical development support, but also to prevent infringement of intellectual property rights.

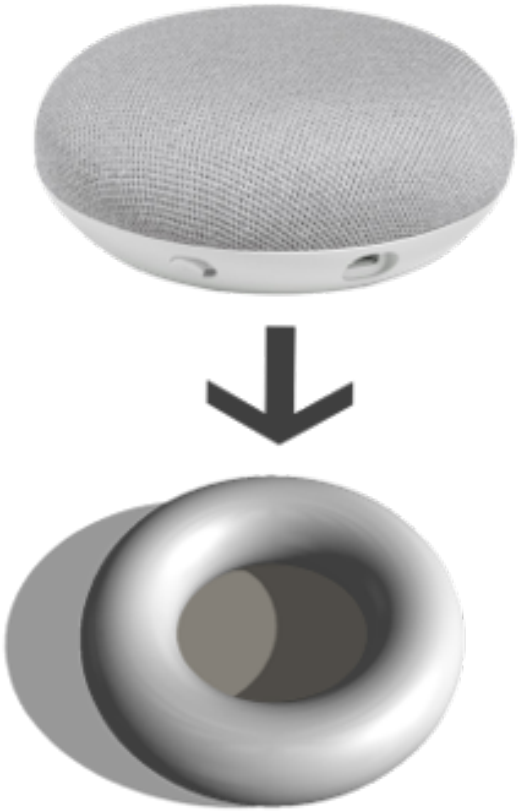

Figure 7: Final designing

  - **Google & Amazon** Our business is dependent on the development of smart speaker. Right now, Google Home and Amazon Alexa combining account for 90% of market share. They are special to our business, our product are against them while we also dependent on their progress. These giants won't stand by and watch us standing in their way, they will have some way to solve this problem. Instead of waiting for competition with these giants, which would be tough for our start-up, we have choice to get in touch with them and seek some kind of coorperation or even been bought in the future.
  - **Speech recognition special company** One of the key value of our product is that we do not hurt the convenience. As long as we need a speech recognition solution which ought to work as better as Google does, it is wise to find a partener who is specialized in this area.

- **Value Proposition**

  - **Keep privacy for our customer** The core value of our product is keeping privacy, to prevent Google and Amazon from tapping ourcustomer and also to make our customer feel secured.
  - **Remain convenience** The most important reason for using a smart sepaker is to make life easier, our product should be a gain for smart speaker in which case we will keep the feature of audio control and fast response.

- **Customer Segments**

  - **People who already have a smart speaker.** This group of people is the so called early adopters who are willing to try fresh things. They are very likely having such trouble. Acoording to the survey [10], 40% of the users of smart speaker do worry about privacy issues. And our product perfectly fit their demands. They are also proven customers who have willing to pay for such advanced products.
  - **People who want a smart speaker but do not for privacy reason** The number of this group of people is considerable, while the willing to pay is hard to say. With the survey [3] conducted

by Business Insider we can say that we have 6.4 million potential customer in this group. The problem here is how to strengthen their will to pay.

- **Channels**
  In the beginning, we will only focus on the online channel. After we start to profit, we will start to touch with physical retailor and promote our product.
  - **Online direct sale** From the financial lecture given by Vittorino, we know that sale through retailor will dramatically increase the cost, which including profit sharing, charges for putting on shelf and transportation etc. Which make it hard to make profit in the beginning. So our plan is to build up our own website first, and open for pre-oder, which help us estimate the interest of potential customers, have provment to ask for investment and increase the profit margin.
  - **Retailor** For long term plan, physical store is necessary. We need chance to display the functionality in real to convince people rather than barely show the demo video. And, also we need physical nodes to support maintenance job. Physical retailor will be a great solution to such problem. However, this distribution channel would ONLY be a supplement for online direct sale. This is not only because of the high costs, but also because most of smart speakers are sold online, like google home through google store and Alexa through Amazon. Online shopping is one of the tendency of our potnetial customer.

- **Competitors**
  - **Privacy Shield** This is a start-up called Privacy Shield. The only start-up who are working on the same problem. Their product is aiming on Amazon Alexa, the most popular smart speaker right now. It is a cover on top of Alexa, generating noise to disturb microphone. From the demo video we can see that their product is kind of bulky. With that cover, Alexa alomost became a colum. And according to the description, this device need to be turn on and off physically by hand, which decrease the usibility and convenience. Even though they are one step ahead of us, their product is not specifically match the market position of us.
  - **Do it yourself** Although can not be count as a product or competiter, this is a very common way for people to solve such problem. People like my parents do not trust such device, they will turn off the speaker when they are not using them. Some people also use tap to seal the microphone, although it doesn't work well. This kind of solution is extremly cheap or even free. But it has the worst user experience.
  - **Google & Amazon** As the manufacturer of smart speaker, it's easy for them to control the usage of data and promise to keep the privacy of customers. However, their nature determines that they have the attendency and motivation to do so. More over, the most important point, customers do not trust them.

  Then comes our solution, our product will not have any internet connection, so there is no chance for us to make us of your data. We are totally trust worthy. And our product will keep the usibility and convenience. Which makes us very competitive amoung them. The figure below domonstrate our superiority among them.

2.2 Business planning

- In the first stage, say 3 months, we will concentrating on developing the first minimal viable prototype. Which will be use to validate our technical solution and demonstrate the viability to potential customer and invester.

- After we got enough feedback, and seed investment we will soon move into next stage, which will spend about 2 month, and finalize at least 20 prototypes which will be used to exhibits in conference and conduct in-field experiment to improve the user experience and collect pre-orders to prepare the mass production.

Figure 8: What is out there

- After that we will settle down the first beta batch and contact manufacturers to massively produce our product. The profits we earn at this stage is expected to be able to support ourself and start to feedback to our investers.

- If everything goes well, it is time for us to plan for the long term. This kind of business is not sustainable just with such small device, especially when your business is depending on others product. We do have the suggestion from Yvo Desmedt that make a homomorphic encrypted smart speaker. This object is too big for us as a start-up, but it is very hopeful for a sustainable business. Which means we can expand our business to the whole LOT industry, to all the smart devices that might encounter privacy problem. We can build our ecosystem and grow strong. With such expectation. we are going to seeking more investment to work on homomorphic remote computation used in speech recognition area.

3 Business development process

We have a detailed roadmap for how to spending money in chart

Figure 9: Cash flow

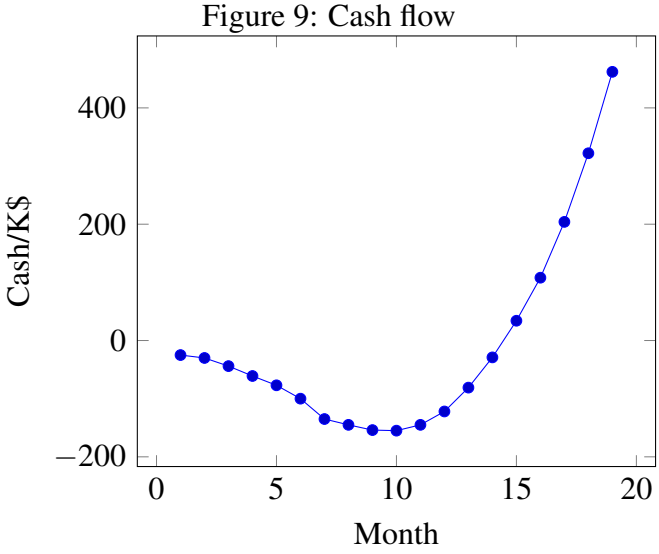

Table 1: Financial detail

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
Volumes						0	500	500	500	800	1200	2000
Unit price						35	35	35	35	35	35	35
Revenues						0.0	17.5	17.5	17.5	28.0	42.0	70.0
marketing												
Website (collect pre-orders)			-4	-2	-1	-1	-1	-1	-1	-5	-2	-2
Communication campaigns			-5	-5	-5	-5	-10	-10	-4	-4	-4	-4
Events						-2			-5			-5
Advertising												
Human resource costs												
Salary for developing MVP	-5	-5	-5									
Building 20 prototypes				-10	-10							
Developing Beta batch						-15	-15					
Long-term development cost								-10	-10	-10	-10	-10
Fixed costs												
Equipment	-10											
Models							-20					
Material	-10					0	-4	-4	-4	-6.4	-9.6	-16
Storage and delivery						0	-2.5	-2.5	-2.5	-4	-6	-10
Cash	-25	-5	-14	-17	-16	-23	-35	-10	-9	-1	10	23
Cumulate cash	-25	-30	-44	-61	-77	-100	-135	-145	-154	-155	-145	-122

4 Self evaluation

References

[1] S. Maheshwari, “Hey, alexa, what can you hear?” https://www.nytimes.com/2018/03/31/business/media/amazon-google-privacy-digital-assistants.html, accessed July 5, 2018.

[2] quartz, “An oregon family’s encounter with amazon alexa exposes the privacy problem of smart home devices,” https://qz.com/1288743/amazon-alexa-echo-spying-on-users-raises-a-data-privacy-problem/, accessed July 5, 2018.

[3] businessinsider, “Consumers are holding off on buying smart-home gadgets thanks to security and privacy fears,” http://www.businessinsider.com/consumers-holding-off-on-smart-home-gadgets-thanks-to-privacy-fears-2017-11?IR=T/, accessed July 5, 2018.

[4] Google, “Data security and privacy on google home,” https://support.google.com/googlehome/answer/7072285?hl=en, accessed July 5, 2018.

[5] M. Morgenstern, “Careless whisper: Does amazon echo send data in silent mode?” https://www.iot-tests.org/2017/06/careless-whisper-does-amazon-echo-send-data-in-silent-mode/, accessed July 5, 2018.

[6] F. Bourse, M. Minelli, M. Minihold, and P. Paillier, “Fast homomorphic evaluation of deep discretized neural networks,” Cryptology ePrint Archive, Report 2017/1114, Tech. Rep., 2017.

[7] B. Li, T. Sainath, A. Narayanan, J. Caroselli, M. Bacchiani, A. Misra, I. Shafran, H. Sak, G. Pundak, K. Chin, K. C. Sim, R. J. Weiss, K. Wilson, E. Variani, C. Kim, O. Siohan, M. Weintraub, E. McDermott, R. Rose, and M. Shannon, "Acoustic modeling for google home," 2017. [Online]. Available: http://www.cs.cmu.edu/~chanwook/MyPapers/b_li_interspeech_2017.pdf

[8] N. Roy, H. Hassanieh, and R. Roy Choudhury, "Backdoor: Making microphones hear inaudible sounds," in *Proceedings of the 15th Annual International Conference on mobile systems, applications, and services*, ser. MobiSys '17. ACM, June 2017. ISBN 9781450349284 pp. 2–14.

[9] U. F. . D. Administration, "Benefits/risks of ultrasound imaging," https://www.fda.gov/radiation-emittingproducts/radiationemittingproductsandprocedures/medicalimaging/ucm115357.htm#benefitsrisks, accessed July 5, 2018.

[10] P. R. Center, "Nearly half of americans use digital voice assistants, mostly on their smartphones," http://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/, accessed July 5, 2018.