Injecting Life into Toys

Songchun Fan, Hyojeong Shin, Romit Roy Choudhury

The Problem Definition:

The paper presents the idea of the fusion of smartphones and toys with the aim to make the toys more interactive and personalized for the children (considering each child is unique). The author has further extended this vision to generalize the application of the paper by focusing on solving the problem of tracking gestures. Basically, the toy detects the children's gestures, cues, and reactions while they interact with it, and then it responds with relevant acoustics and visual response.

Key Idea:

The solution Buzz works on the idea of using the sensors in the smartphones (namely IMU) and signal-processing-based approaches to track the series of gestures (both actual as well as inadvertent) of the children and then determining what the actual gestures are with the help of basic *dynamic time warping algorithm (DTW)*. Then this data is fed to an unsupervised machine learning classifier to learn the gesture patterns over time and then provide the relevant audio/visual output to the children via Toy with the motive to increase the interaction time of the child with the toy. The idea is that as the child interacts more with the toy they try to mimic and repeat the responses to the gesture they performed, which when used healthily helps in the personal development of the child at an early stage. However, the paper only focuses on the first step of recognizing the child's natural gesture patterns, the rest is left for future work.

Important Details:

1. Understanding gestures consist of two parts: Detecting the presence of gestures in the sensor stream and clustering similar gestures for classification.
2. The challenge with detecting the presence of gestures is that we lack any prior data on the relevant children's gestures given all kids perform/act differently. Also, the issue with clustering similar gestures is that the kids have less muscle memory, therefore, they might not repeat the same gesture.
3. DTW (Dynamic Time Warping Algo) is used to compute the dissimilarities between the two gestures.
4. The author avoids using a threshold for detecting the gesture pair but instead uses an unsupervised technique namely, hierarchical clustering to understand and execute the valid gestures from the dissimilarity space.
5. The system has two modules: the front end for detecting potential gestures and generating the acoustic response; Backend to recognize and classify the actual gestures and the training the toy for a relevant response.
6. The paper claims to achieve a precision of 85.1%.

My Thoughts and Criticism:

The paper (in the end) presents a great research topic of recognizing gestures in an unsupervised manner although I believe the use case chosen for this purpose could have been better. The paper's title doesn't quite align with the idea that the paper leads, and I believe there are a lot of limitations and gaps in the research, most of which the author has admitted in the Limitations section of the paper. I felt the paper is not well compiled and deviates from the title as we read further.

However, a few thoughts, questions, and critics came to my mind while reading the papers:

1. Considering the use case in the paper i.e. to inject life in toys, I feel should be more like "Creating Toy Robots" or "Making interactive Toys" or something. But leaving that aside, I do not think we would need a smartphone for this purpose. I mean we have a lot of chips and board computers and independent sensors which will be inexpensive than smartphones and would be easy to maintain over time. Considering the kids aren't usually care about the way they deal with the toys, they can through them off and this might harm the device.
2. The author plans to design the model in a way that aligns the camera of the phone with the toy's eyes and the microphone and speaker with the toy's mouth but that sounds like an impractical design considering toys can be of varied shapes and sizes and no phone has 2 back or 2 front cameras. More work needs to be done here.
3. The design doesn't consider the privacy of the user and the environment sincerely.
4. If not used with guidelines and in a controlled manner (without elder's supervision)- which is one of the aims of this paper, to be able to use to interact with the toy without supervision. This can lead to health concerns in kids and the kid might avoid making friends or playing outdoors and think of the toy as her/his all-time companion. Which can make kids unfit. Of course, the implementation of this solution can be done in a way that we use the concept for some other relevant application under supervision for better results.
5. The solution is tested on an extremely low dataset and lacks the observations in multiple settings. I believe more observations would have brought forward more loopholes and more improvisations could be done.
6. The factor of power and charging the device is completely ignored. This is critical and the author plans to utilize the old smartphones to scale this project with toy ventures. Which I believe can lead to more complex power issues.
7. Adding a smartphone to the toy can increase the weight of the toy and can add a bias to the actions the child might perform with the toy. Thereby leading to a bias in the limited observations made.
8. The author has tuned the smartphone to make a random sound when the kid performs certain actions, which as per the observation encourages the kid to perform the gesture again and this adds to the data for training the model. I believe this step is kind of training the kid to constrain and perform certain gestures again and then eventually learn. This defeats the purpose of the paper "To recognize/track gestures and respond to the kid".