# TEMPORAL-SOUND BASED USER INTERFACE FOR SMART HOME

Kido Tani and Nobuyuki Umezu

Mechanical Systems Engineering, Sci. and Eng., Ibaraki University, Japan

***ABSTRACT***

*We propose a gesture-based interface to control a smart home. Our system replaces existing physical controls with our temporal sound commands using accelerometer. In our preliminary experiments, we recorded the sounds generated by six different gestures (knocking the desk, mouse clicking, and clapping) and converted them into spectrogram images. Classification learning was performed on these images using a CNN. Due to the difference between the microphones used, the classification results are not successful for most of the data. We then recorded acceleration values, instead of sounds, using a smart watch. 5 types of motions were performed in our experiments to execute activity classification on these acceleration data using a machine learning library named Core ML provided by Apple Inc.. These results still have much room to be improved.*

***KEYWORDS***

*Smart Home, Sound Categorizing, IoT, machine learning.*

## 1. INTRODUCTION

IoT devices are widely used for controlling home appliances via Internet. Remote controllers, however, are still the most commonly used technology to operate home appliances. Remote control with Infrared rays has been in use since 1975 in Japan and is a reliable method with less small false detections.

Though remote controllers are commonly used today, there are two major problems. The first problem is that users need to use a particular remote controller corresponding to a specific home appliance. For example, when users try to turn on the air conditioner, they need to find and use a remote controller for the air conditioner. They can't use a remote controller for the TV instead. Such a number of remote controllers are confusing and take up space.

Another problem is that remote controllers are easily left behind without being sanitized. According to a report on COVID19 health crisis (Fig. 1) [1], a large number of viruses were found on remote controllers in a huge cruise ship case. In contrast,COVID19 are detected few from the light switch or the door knob because they might be considered as items which are need to be sanitized.

To solve these problems, we aim to create a new home appliance control device that is easy to manage and users don't need to touch directly. So, we propose a gesture-based interface to control a smart home. Our system replaces existing physical controls with our temporal sound commands.

David C. Wyld et al. (Eds): VLSIA, INWES, CSIP, BMLI - 2021
pp. 67-76, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.112107

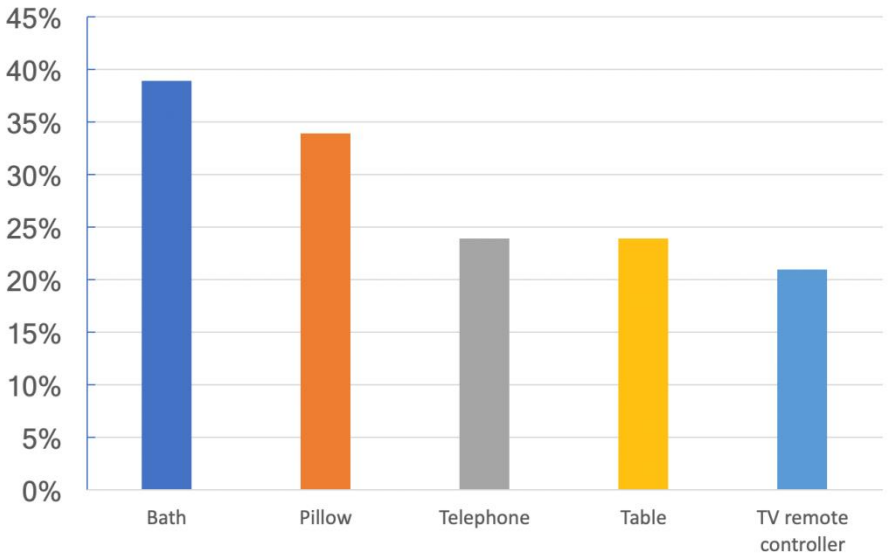


Figure 1. Detection frequency of COVID19 on the Diamond Princess [1]

## 2. RELATED WORK

### 2.1. Sensor array for smart home

An array of sensors has been proposed for recognizing various activities in a home [2]. Their system called SYNTHETIC SENSORS includes multiple sensors such as a thermometer, hygrometer and microphone on its single board design (Fig. 2). When you twist the faucet in the kitchen to get water, various information, such as the sound with twisted faucet with water falling, and the humidity raised by this falling water, would be detected by these sensors. Activity classification is performed on these sensor values.In addition, the use of two SYNTHETIC SENSORS enables more advanced sensing, such as counting how much water flows.

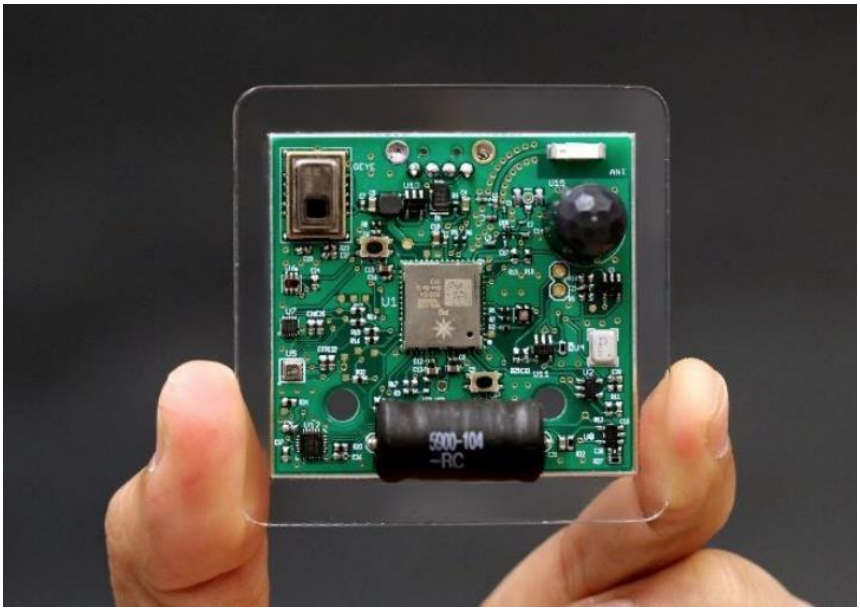

Figure 2. SYNTHETIC SENSORS [2]. Multiple sensors are mounted on a single board.

### 2.2. Controlling with knocking

A small IoT device called “Knocki” shown in Fig. 3 can be mounted on a surface such as a wall [3]. By knocking the wall around this Knocki for a specific number of times, users can turn on or send messages to linked appliances such as lights, TVs, and air conditioners. This device can be used in noisy environments because it does not relay on a microphone.


Figure 3. Knocki attached to a wall [3].

## 3. TEMPORAL-SOUND BASED USER INTERFACE

### 3.1. Overview of the proposed system

In our method, we obtain signals from an accelerometer attached on the user’s wrist to record its vibration. If the recorded signals match one of pre-defined commands for home appliance operation, the system sends a corresponding infrared command to control that specified appliance. We use simple gestures, such as knocking on or scratching the table to control home products.

The processing flow of the proposed system is shown in Fig 4.

1: The user wears the device,
2: performs a specific action to generate vibration.
3: The system recognizes the action from sensor information,
4: requests the remote controller to control appliances.
5: Home appliances react to that command.

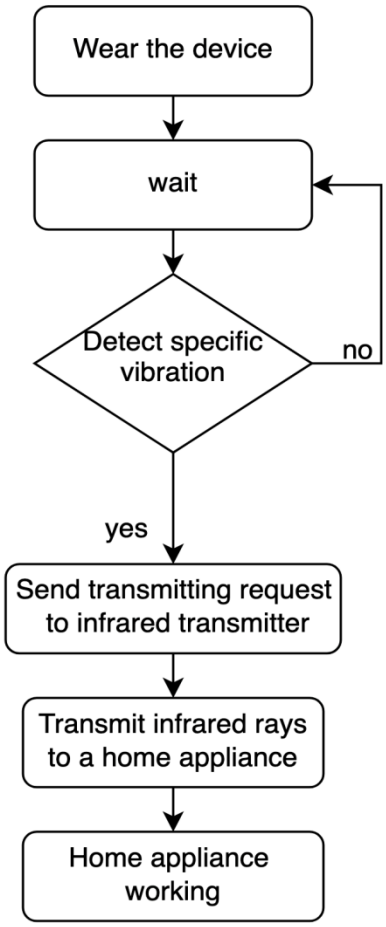


Figure 4. The flow of the system

## 3.2. Recording Temporal-sounds

For a preliminary experiment, we developed a prototype system that has a sound sensor and microphone, instead of an accelerometer. User gesture for our prototype are recognized not with the vibration generated by a gesture, but with the sound by that gesture. We collected a number of sound recordings used for learning gesture classification. These sounds were generated on a wooden desk and recorded with a USB microphone at a distance of 3 (cm) from the desk surface. We recorded 150 sound files for each of the following 6 types: 1) single mouse click, 2) double mouse click, 3) single knock on the desk, 4) double knock on the desk, 5) triple knock on the desk, and 6) double claps. Every waveform image shows a characteristic waveform (Fig. 5). Although mouse clicks are not used as the operation command at last, we add them as a test to confirm the operation of the system.

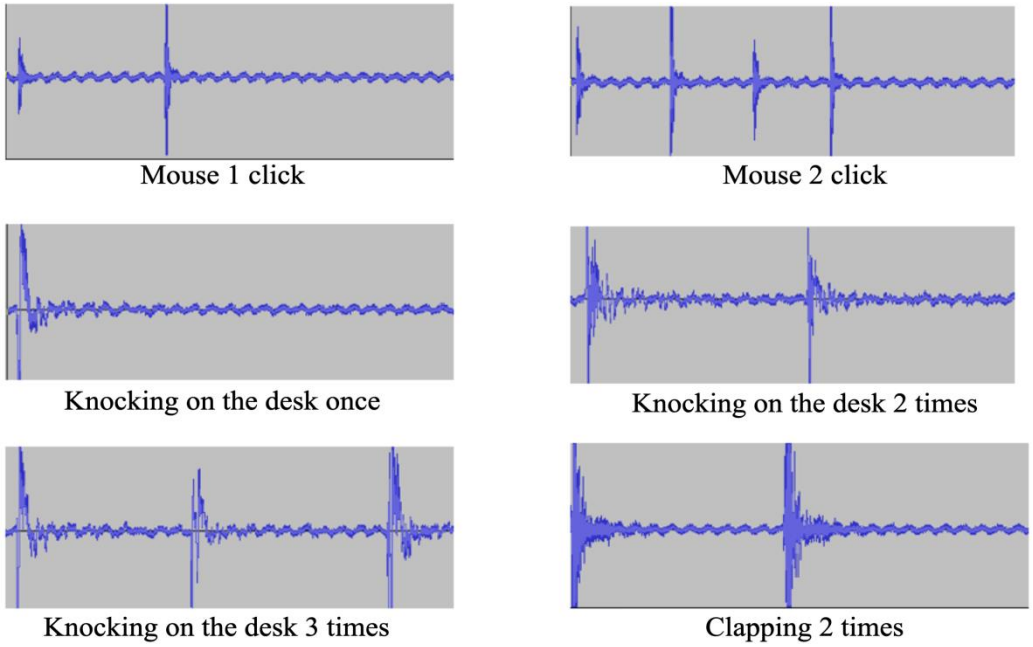


Figure 5. Wave images of 6 types of sounds

### 3.3. Converting to Spectrogram

Because it is difficult to perform deep learning with one dimensional sound signals, we converted them into two-dimensional spectrogram images [4]. A spectrogram is an image that represents the amplitude, frequency, the time of a sound by performing frequency analysis and using color shading. In the example of Fig.6, the closer the color of each point is to red, the higher the amplitude of the frequency.

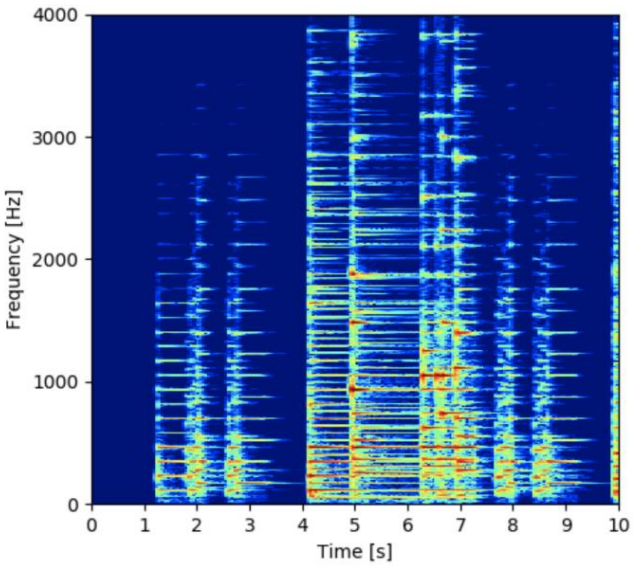


Figure 6. Example of the spectrogram

### 3.4. Learning spectrogram

We converted 150 sounds for each of 6 types into 900 images (Fig.7). We used a CNN model for learning spectrogram. Neither data enhancement of the training image nor drop-out is used. 30 of 150 spectrogram images for each sound are used as test data to improve accuracy. The training time was approximately 5 hours performed on a PC with AMD Ryzen 3900x, 32GB main memory, GeForce RTX2070 SUPER 8GB.

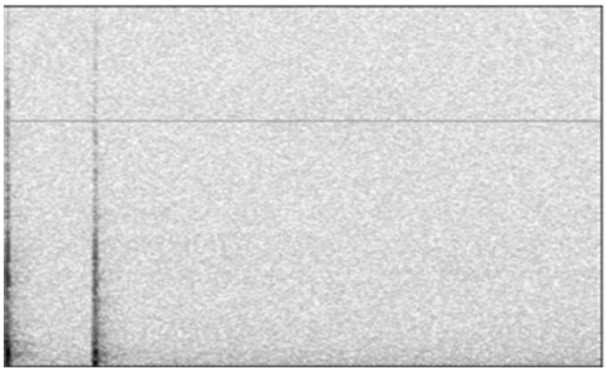

Figure 7. Converted image of spectrogram

### 3.5. Developing a remote controller

We developed a remote controller using Raspberry Pi [5], a popular System on a Chip (SoC) computer system. We implemented learning processes with infrared (IR) rays from other controllers and an IR transmission circuit on this small controller (Fig. 8). Our learning process with IR signals is based on a library of “IR Record and Playback” by pigpio.

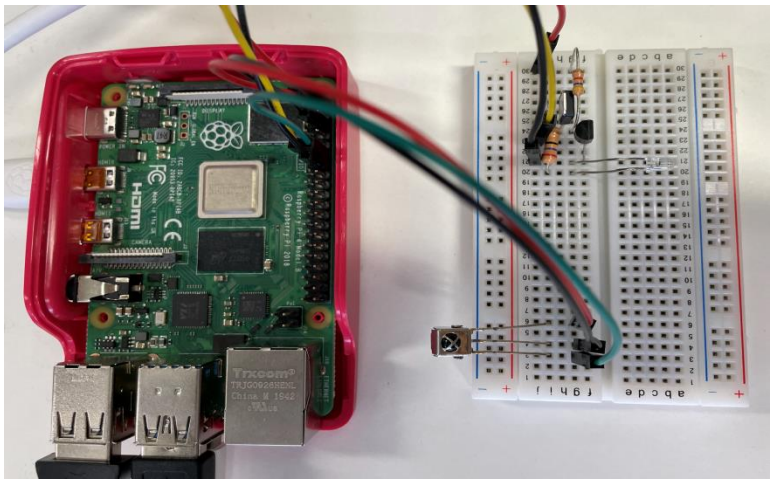

Figure 8. Remote controller using Raspberry Pi

### 3.6. Acquisition of acceleration data

In this experiment, we collected acceleration data using Apple Watch Series3[6]. We use CMMotionManager [7] to acquire motion data, which can acquire the motion sensor data in the device published by Apple, when the button displayed in the Apple Watch was tapped, the program acquires motion data including acceleration data. Five types of gestures were collected

(resting, knocking twice on the desk, knocking three times, clapping twice and knocking with a thump-thump thump rhythm).We recorded about 100 times for each gesture. The gesture occurred three times in one data. Figure 9 shows a graph of the transition of acceleration for two knocks.

It can be seen that the X-axis and Z-axis reacted twice.

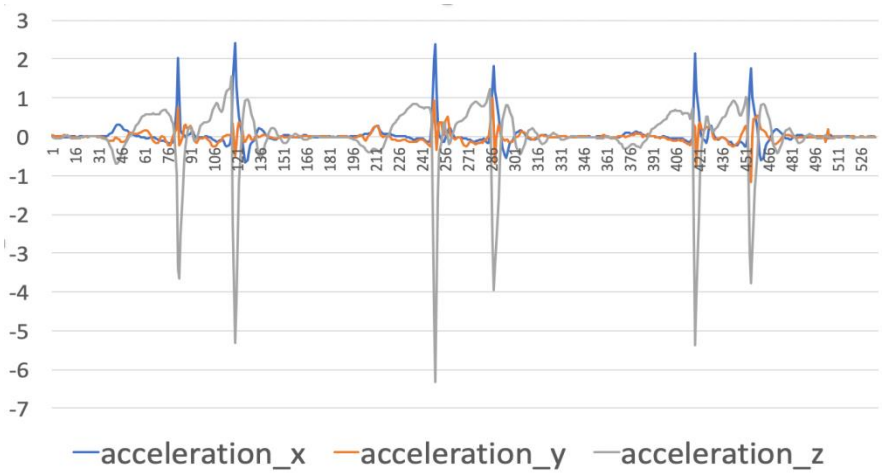


Figure 9. Acceleration for knocking two times on the desk

Core ML [8], provided by Apple Inc., is used to train the collected acceleration data.It can perform machine learning on-device by simply preparing the specified data and there is no need to build a learning layer or set up an activation function. As shown in Fig. 10, it supports various types of learning. In this research, we use activity identification. In this research, we use Activity Classification.

## 3.7. Learning acceleration data

We used Core ML's activity identification to learn. There are 16 accelerations that can be used for learning, acceleration_x,acceleration_y, acceleration_z, attitude_pitch, attitude_roll, attitude_yaw, gravity_x, gravity_y, gravity_z, quaternion_w, quaternion_x, quaternion_y, quaternion_z, rotation_x, rotation_y, rotation_z. In this study, we use only acceleration_x, acceleration_y, and acceleration_z because there is no significant difference between using all of acceleration data and using only 3 accelerations.

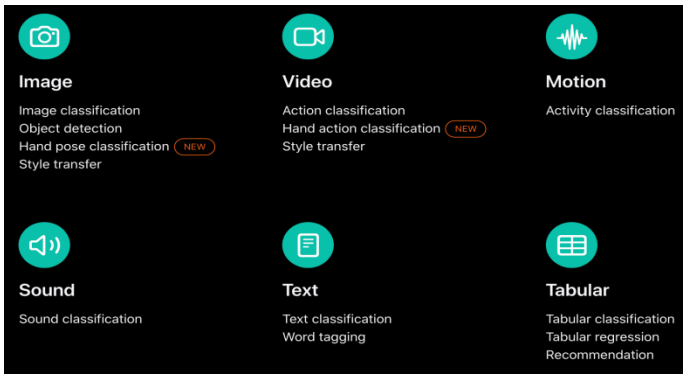


Figure 10. Available model for Core ML framework from Apple Inc. [7]

# 4. CLASSIFICATION EXPERIMENTS

## 4.1. Classification experiment using Raspberry Pi

We conducted a series of experiments to evaluate recognition results based on our CNN learning. These experiments were performed using a microphone connected to the Raspberry Pi. We made 20 trials for each sound to rate the recognition precision of the proposed method.

The experiment results are shown in Table 1. Sounds of clicking on mouse once and twice are not included in Table 1 because these two sounds are too quiet to detect.

All 4 sounds have poor accuracy. It might be we because we used sounds recorded using microphone attached to a computer, not a Raspberry Pi when we train the model.

Table 1. Result of the experiment using Raspberry Pi. Columns shows the input. Rows shows the result of the classification.

| result<br><br>input | Knocking on the desk (once) | Knocking on the desk (twice) | Knocking on the desk (three times) | Clapping (Twice) |
|---|---|---|---|---|
| Knocking on the desk(once) | 40% | 60% | 0% | 0% |
| Knocking on the desk(twice) | 0% | 85% | 15% | 0% |
| Knocking on the desk (three times) | 0% | 40% | 60% | 0% |
| Clapping (Twice) | 10% | 25% | 10% | 55% |

## 4.2. Classification experiment using Apple Watch

We conducted experiments to measure the classification accuracy of Core ML model. We set the parameters of Core ML: prediction windows size (130), sample rate (50). After wearing the Apple Watch (Fig.11), we made 20 trials for each gesture to rate the recognition precision of the proposed method. Table 2 shows the result of the experiments. Almost all the trials are classified as the knocking on the desk 3 times. This might be practically because the duration of knocking on the desk 3 times surpasses the time window used for the recognition process, which requires further investigation.


Figure 11. Running the classification program. The Japanese sentence in the screen means motion classification.

Table 2. Result of the experiment using Apple Watch. Columns shows the input. Rows shows the result of the classification.

| result / input | Neutral | Knocking on the desk (twice) | Knocking on the desk (three times) | Knocking on the desk (2-1) | Clapping (Twice) |
|---|---|---|---|---|---|
| Neutral | 0% | 0% | 100% | 0% | 0% |
| Knocking on the desk (twice) | 0% | 0% | 100% | 0% | 0% |
| Knocking on the desk (three times) | 0% | 0% | 100% | 0% | 0% |
| Knocking on the desk (2-1) | 0% | 5% | 95% | 0% | 0% |
| Clapping (Twice) | 0% | 5% | 95% | 0% | 0% |

## 5. CONCLUSION

Although the remote controllers are commonly used in daily life, they have two major problems. The first problem is that users need to use a particular remote controller corresponding to a specific home appliance. A number of remote controllers are confusing and take up space. Another problem is that remote controllers are easily left behind without being sanitized. According to a report on COVID19 health crisis, a large number of viruses were found on remote controllers in a huge cruise ship case. To solve these problems, we proposed temporal-sound based user interface that is easy to manage and don't need to touch directly.

For a preliminary experiment, we developed a prototype system that has a sound sensor and microphone. After recoding command sounds for controlling home appliances, we converted them into spectrogram and conduct CNN learning using the images. We recorded the acceleration with Apple Watch and we conduct activity classification for the data using Core ML.

To evaluate the CNN model, we conduct the classification experiment using Raspberry Pi. The accuracy is poor for the most part. We also conduct experiments to measure the accuracy rate of Core ML model using Apple Watch. Almost all the results are classified as knocking on the desk 3 times. Both models have much room to be improved.

## 6. FUTURE WORK

In our experiments, we used spectrogram images recorded with a microphone connected to a PC for learning, and those images with Raspberry Pi for recognition tasks. The recognition accuracy might be improved if we use spectrogram images recorded Raspberry's microphone for both learning and recognition. We are conducting experiments with our recognition model using sound files with the microphone connected to Raspberry Pi.

Alexa [9] is one of the most popular speech recognition interfaces and a number of appliances have compatibility [10] with it. Our approach could solve some of disadvantages in speech interfaces such as longer voice commands, and effects by background noises.

To replace existing remote controllers with our system, more operation commands are needed, such as turning up/down the volume, changing the channel. We are planning to increase and regulate the gestures to correspond more commands.

We are planning to conduct user experiments and questionnaire to measure the quality of our system. First, the participants in the experiment sit in a chair and type the instructed text into the computer. While they are typing the text, they are instructed to turn on/off the light four times at random time points. We measure and compare the time between the instruction and actual operation when participants use the remote controller or our device. The questionnaire is based on the following items: ease of movement, ease of understanding gesture commands, interest in using it on a daily life, differences from the remote controllers and comment. We collect answers to the questions on a 5-point scale from 1: dissatisfied to 5: satisfied.

## REFERENCES

[1] National Institute Of Infection Disease , “Report on the Diamond Princess Environmental Inspection” , https://www.niid.go.jp/niid/ja/diseases/ka/corona-virus/2019- ncov/2484-idsc/9849-covid19-19-2.html, (Accessed 2021.5.1)

[2] GIERAD LAPUT , “SYNTHETIC SENSORS” , https://www.gierad . com/projects/supersensor/%3E, (Accessed 2020.10.20)

[3] Knocki, "Knocki" , https://knocki.com/, (Accessed 2020.11.13)

[4] MathWorks , " ド キ ュ メ ン テ ー シ ョ ン Spectrogram" , https://jp.mathworks.com/help/signal/ref/spectrogram.html, (Accessed 2021.7.6)

[5] Raspberrypi, " Raspberry Pi 4", https://www.raspberrypi.org/products/raspberry-pi-4-model-b/ , (Accessed 2021.7.6)

[6] Apple, “Watch –Apple”, https://www.apple.com/jp/watch/, (Accessed 2021.11.25)

[7] Apple, “CMMotionManager”, https://developer.apple.com/documentation/coremotion/cmmotionmanager, (Accessed 2021.12.8)

[8] Apple, “Create ML”, https://developer.apple.com/machine-learning/create-ml/, (Accessed 2021.11.25)

[9] Amazon.com, “Alexa”, https://www.amazon.com/b?ie=UTF8&node=21576558011, (Accessed 2021.12.6)

[10] Hamilton Beach, “Smart 12 Cup Coffee Maker – Works with Alexa® certified”, https://hamiltonbeach.com/smart12-cup-coffee-maker-works-with-alexa-49350, (Accessed 2020.11.13)