February 27, 2019

Display Orientation User Experience (UX) Enhancement Through a Sensor Array

ChiLin Kuo

Chin-Kuo Huang

Display Orientation User Experience (UX) Enhancement Through a Sensor Array

Abstract:

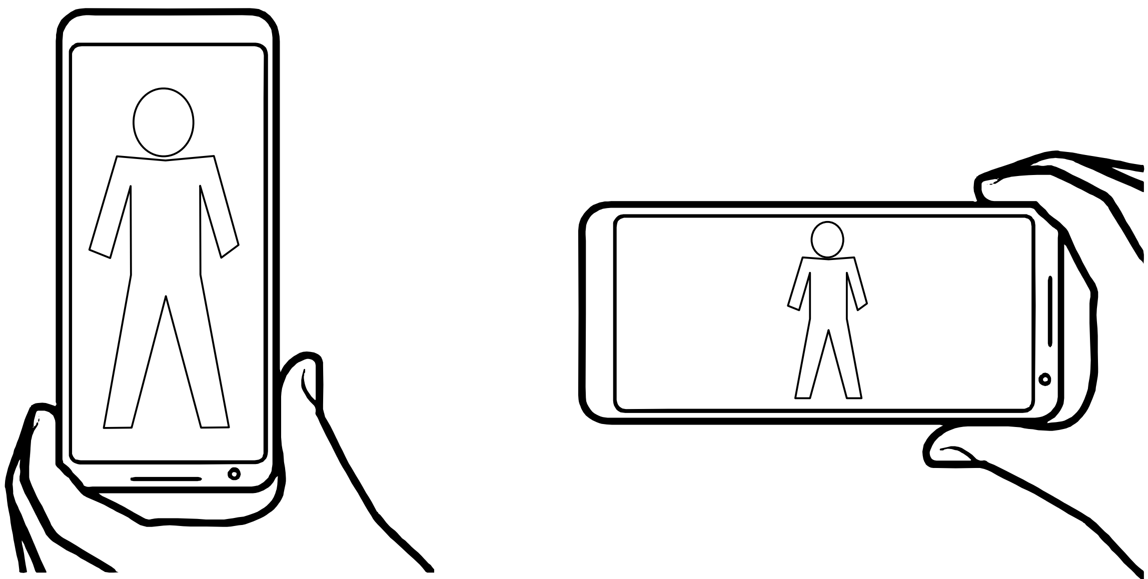
This publication describes a computing device that captures a user-initiated sound by incorporating a microphone array. The computing device analyzes the captured sound from each microphone by using a model capable of determining a variation in the sound power of the captured sound signal by each microphone. To conserve the resources of the computing device, machine learning trains the model remotely or off-line. The ability of the computing device to determine the user-initiated sound power variation captured by each microphone enables the device to determine the location of the user in respect to the device and, thus, allows the device to respond accordingly (e.g., auto-rotate the screen towards to the user).

Keywords:

Voice command, user-initiated sound, sound command, vocal command, sound power, sensor array, microphone array, user experience (UX), display orientation, machine learning (ML), neural network, deep learning, artificial intelligence (AI).

Background:

Various computing devices (e.g., smartphone, tablet, digital camera) and the accompanying operating systems (OS) have an "auto-rotate" capability based on a device's screen orientation, as shown in Figure 1. In the following discussion, operations in a smartphone illustrate the features of this disclosure.



a) Smartphone held vertically **b) Smartphone held horizontally**

Figure 1

Figure 1 shows the auto-rotate capability that one may have already observed while using a smartphone. As shown in Figure 1a, when a user holds the smartphone vertically, the displayed image also rotates vertically. But, when the user holds the smartphone horizontally, as shown in Figure 1b, the displayed image will rotate 90 degrees. This smartphone feature aids the user to view the image without tilting his or her head. To accomplish this auto-rotate feature, designers may employ various means, such as accelerometers or magnetometers.

Nevertheless, there are cases when this auto-rotate feature does not work effectively. Figure 2 helps demonstrate such a case.

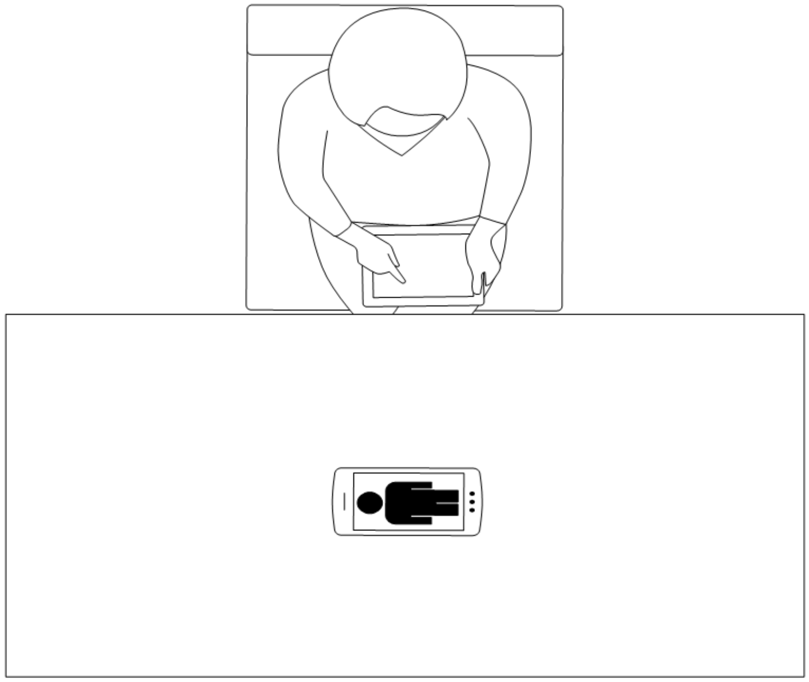


Figure 2

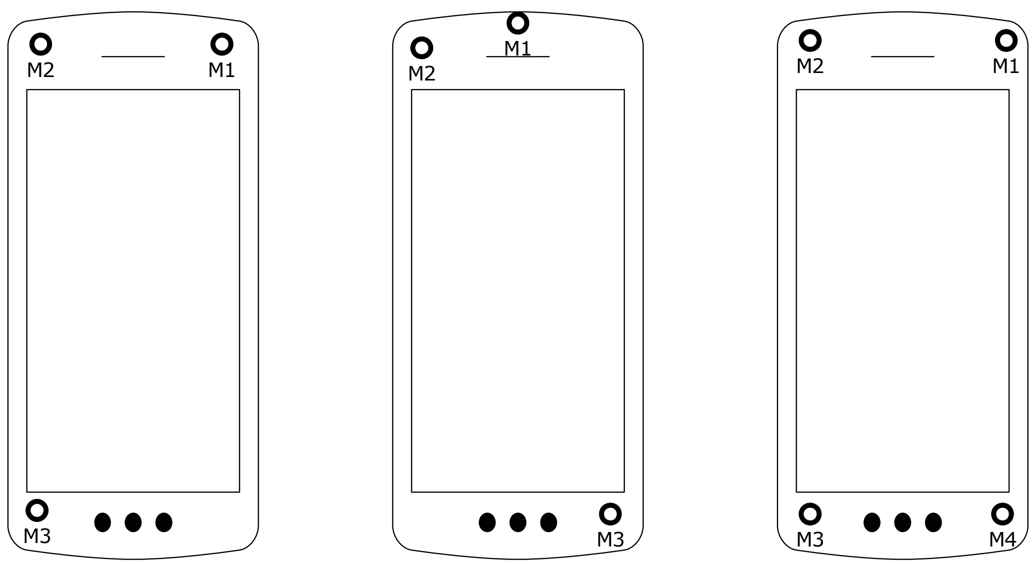
In Figure 2, consider Jane working on her work notebook. She receives a photo message on her private smartphone from her son. The smartphone is on the table, and the only way for Jane to view the private message is to stop working, reach for the phone, physically pick-up the phone, wait for the message to auto-rotate, and then view the message.

As this example illustrates, it is desirable for a computing device (e.g., smartphone, tablet) to have the capability to auto-rotate the displayed message even in instances when the device is laid on a flat surface without the user needing to touch the device. Furthermore, it is desirable for the computing device to determine the user's location in respect to the display.

Description:

This publication describes a display orientation system of a computing device equipped with a multi-microphone sensor array. The computing device determines a user's location by analyzing a user-initiated sound. This allows the device to enhance a user's experience (UX) by responding accordingly (e.g., auto-rotate the display towards the user) without the user needing to touch the device.

The computing device uses three or more non-collinear microphones embedded on the device, which detect the same sound independently. Figure 3 shows examples of these microphone placements on the computing device.



a) An example of three embedded microphones **b) Another example of three embedded microphones** **c) An example of four embedded microphones**

Figure 3

Figure 3 shows three examples on the possible locations where the microphones (labeled M1, M2, M3, and M4) may be embedded on a smartphone. Nevertheless, the invention may apply to any computing device where determining the user location may be desirable. Figures 3a, 3b, and 3c also show that the microphones are not embedded collinearly. Because two microphones can be collinear and equidistant from the user, the computing device uses three or more embedded microphones. Theoretically, however, many more microphones can be used.

A user-initiated sound may be a user's vocalized phrase (e.g., OK phone, over here OS), clapping (e.g., one clap, two claps, a specific clapping rhythm), finger snapping, whistling, etc. As the user-initiated sound signal propagates through a medium, such as air, each microphone detects a different sound power level or sound strength, because the farther the distance from the sound source (e.g., the user's vocal chords), the weaker the sound power at the destination (microphone). As one can imagine, the sound power variation to each microphone is small given that the embedded microphones are clustered close to each other, while the user-initiated sound source is relatively far to each of the microphones. The user may be one meter away from the device, while all the microphones may be clustered within ten centimeters of each-other. To this end, to determine such small variations in sound power, this invention leverages machine learning (ML), as shown in Figure 4.

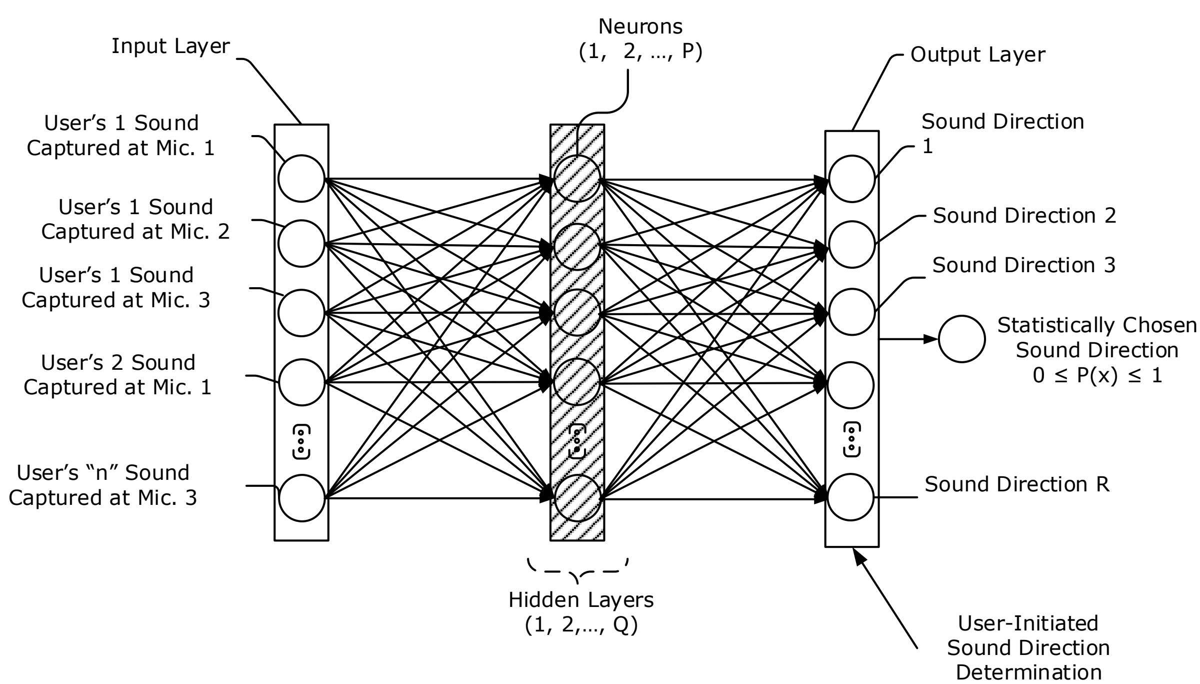


Figure 4

Figure 4 demonstrates a neural network used for a computing device with three embedded microphones, however, a similar neural network may be used for devices with more than three embedded microphones. The neural network in Figure 4 illustrates an input layer, several hidden layers, and an output layer. The input layer includes user-initiated sounds from "n" number of users, which are captured by three microphones. There are "Q" number of hidden layers with up to "P" number of neurons in each layer. There can be a different quantity of neurons in each hidden layer. The output layer includes "R" number of bins with different probabilities on the direction of the user-initiated sound source. The bin with the closest probability to one (1) is interpreted as the correct output.

Given the large computational power that machine learning can use to train a model to distinguish such small sound power variations from a user-initiated sound source, the model training can be performed on a cloud, server, or other capable computing device or system.

Periodic model updates are sent to each user's computing device, which allows the user's computing device to execute the model even if that device does not have the resources to update the model itself.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., a user's preferences or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

Figure 5 further demonstrates how the user-initiated sound triggers the auto-rotate feature.

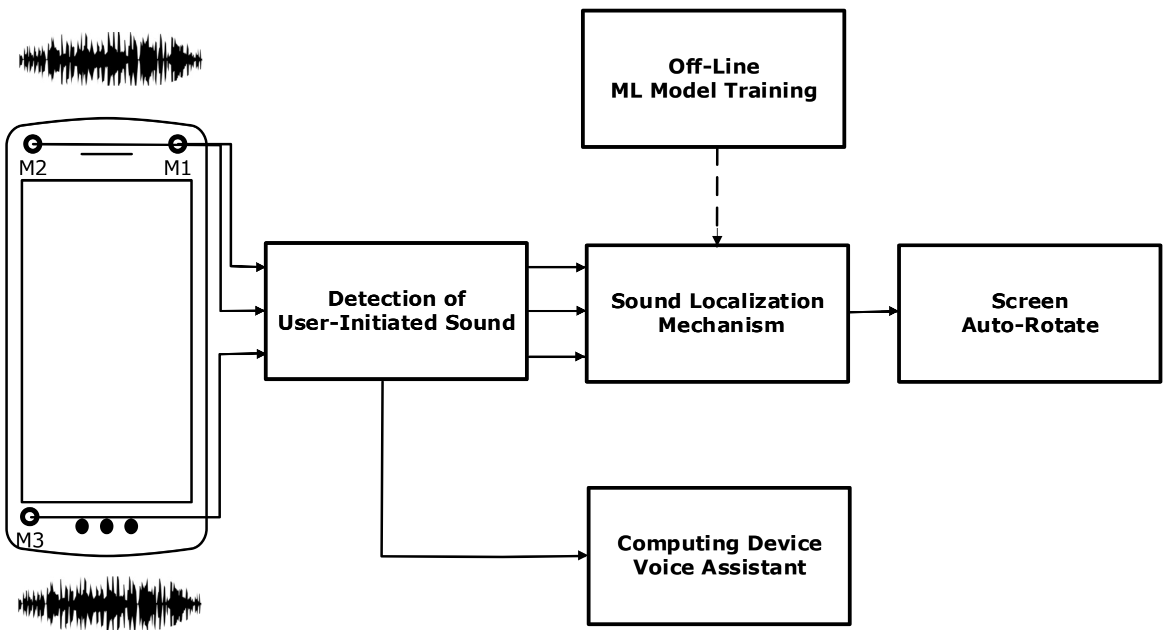


Figure 5

A computing device with three imbedded microphones, as shown in Figure. 5, detects the user-initiated sound (e.g., the user says: "OK phone"). The sound-localization mechanism analyzes the captured user-initiated sound signal by each microphone. The sound-localization mechanism incorporates the capability to measure the sound power and employs the off-line or remotely trained model, which determines the sound power variation from each microphone. The microphone that captures the highest sound power is located closest to the user (e.g., user-initiated sound source). After the sound-localization mechanism determines the user's location, the screen auto-rotates accordingly. In the case when there is uncertainty on the location of the user, the computing device, through a voice assistant, may ask the user to initiate the sound once more or may ask other questions that may help the computing device determine the user's location. Later, these uncertainties may be used as inputs to the machine learning model to improve the model at future updates.

Figure 5 illustrates an exemplary use of this invention. Nevertheless, the ability of a computing device to determine a user's location based on a user-initiated sound may have other uses, such as automatically moving (e.g., auto-rotating) a mounted light fixture, a digital camera, a sound speaker, a physical display (e.g., a desktop monitor or television), and the like.

Although several of the examples presented in this publication describe machine-learning techniques that use vocal sound detected by microphone sensors, any number of other sensors may be used to implement the techniques described herein. For example, a radar-based sensor array (e.g., radar transceivers) may be used to detect a user-initiated movement like a hand wave or nodding. Radar-based sensor arrays can correctly interpret the location of the user and auto-rotate a screen display orientation toward a user.

In summary, the computing device captures the user-initiated sound using a microphone array. The device analyzes the captured sound signal by each microphone and uses a model that determines the variation on the sound power captured by each microphone. The model can be trained off-line using machine learning to iteratively improve the quality of the model with each update. The detection of sound power variation in each microphone enables the computing device to locate the user and respond accordingly, such as to auto-rotate a screen's display orientation toward a user.