TECHNISCHER LÖSUNGSVORSCHLAG   Aktenzeichen: A2017-009

Titel: **„Voice Control Device Selection"**

Autor/en: **Matthew Schwind**

Unternehmenseinheit: **Diehl AKO Stiftung & Co. KG**

Description:

Voice control device selection through ranging

The system determines which voice controlled device should respond to a user. This is determined by proximity of the user to the device. The system is decentralized where each device shares its data with all other devices and compares the data of all other devices.

A user gives a voice command in range of multiple voice control devices. All of the devices in range will respond, and this can disrupt the interaction of the user with an individual device. This also creates wasted bandwidth (sending the voice command), compute cycles (processing of the voice command) and energy. For an individual this is negligible loss. However, for the company processing the commands of millions of users, this waste could become significant. This problem will grow with the increasing adoption of voice control systems.

Voice controlled technology is relatively new. The patents below have in detail a similar system however their system involves a centralized system. In this systems all the devices record the audio command of the user and upload the data with a processing request to a central server. Each request creates a session on the server to process the data, determine the desired command and derive a response. If the system determins that multiple sessions share the same user request, then one of the sessions will be selected for response. While all other sessions are cancelled.

A user says a command that is picked up by several devices. This causes the devices to enter the wake state. Each device that is in wake state recordes the audio of the user

speaking the command. Then the device calculates its range to the user by analyzing the recorded audio. The devices then share their range with all the other devices that are in the wake state. The devices will compare their range to the ranges of the other devices. If a device does not have the closest range, then it will return to idle. The device with the closest range will continue interacting with the user, by sending a request to process the audio data from the user. This also means that a continued session, or conversation with a device, will not involve any other devices in range. Until the session with the interacting device is ended.

The suggested system is decentralized. This means that the processing is handled faster and with no waste for the company or the user.

For the centralized system detailed in part two. Each extra device in range of the user creates waste and loss. For example, if two devices are in range of a user voice request. Two requests (and packets of data) are sent to the central server – one request is waste. Both create sessions and are processed until it is determined that one is redundant. One session is wasting computer resources and power. The algorithm that is running on all sessions trying to determine if there is redundancy is also wasting resources and power. This waste is multiplied by each extra device in range of the user. If the request requires multiple back and forth interactions then this waste is multiplied by each interaction.

The decentralized system has none of this waste.

- ❖ Problem: If there are several voice enabled devices in one area, which device answers when the user interacts with the device.
  - ➢ Current voice controlled device start an interaction with a user through the use of a wake word. For the amazon echo and amazons voice assistant Alexa the wake word is "Alexa".
  - ➢ If there are more than one Alexa enabled devices within hearing distance of a user. Then saying the wake word "Alexa" will cause all the devices to wake up, respond to the user and begin recording.
  - ➢ All devices do not respond in the same way and at the same time. Some turn on a light, some chime, and other give a voice response before they begin listening to the user.
  - ➢ Best case scenario you have two devices responding and giving the same information. Making this only is an annoyance to the user. But the service provider will be paying for processing the data sent by both devices.

- ➢ However, more likely is the response of one device being recorded by the other(s) and interfering with the interaction of the user. As well as increasing the cost of processing the user's interaction.
- ➢ This will become more of a problem as the use of such devices increases.

- ❖ Solution: Range of the user, share the ranging data with the other devices and only the closest device responds (see drawing).
  - ➢ When the user says the wake word all the devices that heard the wake word record the range of the user. Before responding to the user the devices share their range to the user with the other devices. Each device will compare its range to the ranges of all the other devices. Only the device with the lowest range will respond to the user for the duration of the interaction.
  - ➢ The default state of the devices is the "Idle state". In which they are listening for the wake word.
  - ➢ Once the device hears the wake word it moves into the "Wake state". In this state each device will calculate the range of the user and share it with the other devices.
  - ➢ Ranging of the user can be done by analyzing the volume of the wake word by each device, and determining the dB level.
  - ➢ Another possibility is through recording the time at which the wake word was detected (more difficult)
  - ➢ Then the devices share their range info with all other devices that are in the "Wake state".
  - ➢ This can be done over WIFI or through what will be described later on as "High Frequency Audio Communication". If over WIFI, the devices need to all be on the same network, and they will be sharing their ranging data with all devices on the network
  - ➢ The devices will compare their ranging data against the other devices data. If theirs is not the closest (i.e. has the highest dB level or lowest time stamp), then the device will return to the Idle mode state.
  - ➢ The device with the closest range to the user will switch to the "Interaction state". This is the final state where the device will interact with the user.

Drawings: