Auto Group - Automatically grouping speakers based on sound

Authors: Alexander Ryde & Alexander Green

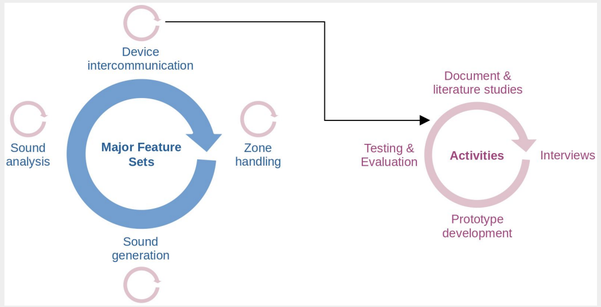
Problem

Axis expressed a goal of further improving their existing manual grouping functionality in a site of network speakers. Grouping a few speakers may not require too much work, however in a situation with 200 speakers it can take a long time and be perceived as complicated. An automatic grouping process based on sound was to be investigated.

Method

The chosen method was an iterative workflow with returning focus on major feature sets. Those feature sets were device intercommunication, zone handling, sound generation and sound analysis.

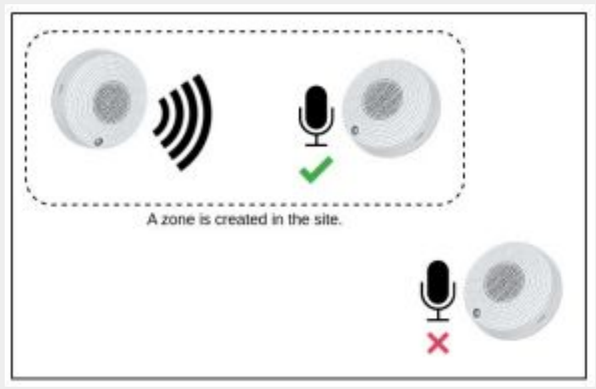
During work on each major feature set there were smaller iterations within focused on activity types; document studies, literature studies, interviews, prototype development and finally testing and evaluating the prototype.



The main motivation for this was for the benefit of focusing an activity, such as document & literature studies, on a smaller amount of relevant technologies at a time.

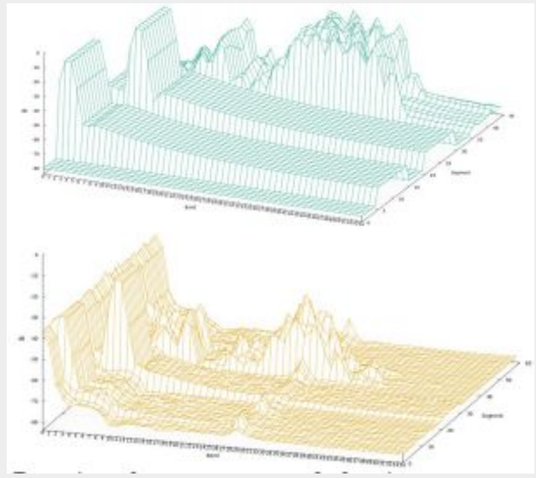
Solution

Through the use of a new API endpoint a speaker can be instructed to play a predetermined configuration sound. Other speakers in the same site will at the same time be instructed to start recording. The recording speakers that detect the configuration sound gets added to the playing speaker's group.



Result

Devices communicate using MQTT messages to send instructions to record/play or to report successful detections of the configuration sound. GStreamer is used for the actual sound playback and sound recording as well as in the sound analysis. The sound from a playing device consists of three different sections separated by silence. The first two sections are sine-wave tones at 600 Hz and 2200 Hz while the last section is a wider high-frequency sound focused roughly around 11000 Hz.
A listening speaker will analyze its recorded sound and report magnitudes, for three frequency bands, that are above the background noise level. The site leader device acts on the reports by adding ungrouped listening devices to a player device that was heard. There is also an option to start a longer process of letting all speakers located in the site to play in sequence, automatically grouping all devices in the site.



Discussion

The stated goal was that the prototype should group three speakers, in one room, into one group and two other speakers, in another room with a closed door in between, should end up in a separate group. The prototype functioned well in such a situation.

Through evaluation of the prototype, concerns regarding noise, frequency spread and latency were identified and illustrated using data output from the prototype. While the initial goal for the prototype was that it should work in work in quiet surroundings, the prototype handles constant background noise as well as sudden noises shorter than 0.8 seconds. Device latency did not cause a problem during testing of the prototype but high enough latency would cause a device not to be grouped when it otherwise would be. Solutions for the problem were identified and described along with other future development ideas as part of the thesis report.

In conclusion the results showcase that automatically grouping speakers based on sound is a viable strategy.