Technical Disclosure Commons

Defensive Publications Series

October 2020

Audio Selftest For Automated Fault Detection In Meeting Room Hardware

Anonymous

Audio Selftest For Automated Fault Detection In Meeting Room Hardware

ABSTRACT

Meeting rooms are important for team collaboration and tend to have a high occupancy rate. It is important to detect and repair malfunction of meeting room hardware that is used to conduct video conferences. This disclosure describes audio selftest techniques that can automatically detect faults and enable preemptively fixing the detected faults to ensure that each meeting room meets a minimum set of quality standards. The audio selftest includes playing back sounds through in-room speakers and detecting the sounds via in-room microphones. Common problems such as disconnected or broken speakers/ microphones, swapped microphones or microphone arrays, miswired microphones, etc. can be diagnosed. Further, interference in the room can be detected and rooms can be classified as suitable or unsuitable for meetings based on ambient noise detected in a silent room by classifying the detected signal. Such automated diagnosis can enable frequent testing of meeting room hardware, provide early detection of malfunctioning equipment, and enable faster repair.

KEYWORDS

- Audio selftest
- Audio diagnostics
- Conference room
- Meeting room
- Video conferencing hardware
- Fault detection
- Microphone array
- Microphone array
- Noise floor

BACKGROUND

Many organizations have office facilities that have a large number of meeting rooms with video conferencing hardware that are used by the organization's staff on a regular basis. Such rooms are important for team collaboration and tend to have a high occupancy rate. The equipment that supports video conferences in such rooms can malfunction and it is important to detect and correct problems as soon as possible. However, given the number of rooms and the time/effort required for manual verification, it is impractical to perform periodic preventive maintenance on meeting room hardware.

DESCRIPTION

This disclosure describes audio selftest techniques that can automatically detect faults and enable preemptively fixing the detected faults to ensure that each meeting room meets a minimum set of quality standards. Faults can appear at different points during the lifetime of the meeting room audio hardware. During system installation, for example, faults can be introduced such as the unintended swapping of unlabeled or mislabeled cables. In another example, faulty or incorrectly fitted connectors can appear to be properly connected upon visual inspection but actually fail to make electrical contact.

Automation enables a reduction in the staff required to maintain a fleet of video conferencing units. The audio selftest techniques described herein provide a powerful tool to validate the current health status of equipment in a room at the time of initial provisioning, after a repair, and throughout the rest of the equipment lifetime. Failures are unavoidable, but a self-aware system as described herein can reduce the mean time to repair (MTTR) and increase overall system availability.

Automatically detecting faults in the room's audio system has a big impact for several reasons:

1. There may be a significant range between equipment in a room being properly configured and being completely unusable. For example, if only one of several installed microphones in a room is functional, the room is still somewhat usable, although:
    - audio quality will likely be degraded
    - audio from the room may be more susceptible to noise and interference
    - speech from individuals who sit far away from the sole working microphone may not be adequately heard.

   In general, a repair that is applied after detection of perhaps a single microphone failure is better than attempting repair when the last microphone in a room fails.
2. User complaints often do not include sufficient information about the state of the equipment in a room. While end users can diagnose faults and observe degraded quality, they are often unable to provide sufficient information regarding the cause of the failure. For example, in response to user feedback that a room "sounds weird," substantial effort may be required to diagnose and root cause the problem. A service technician first has to reproduce the state in which the failure occurred, e.g., the "weird" sounds might only be audible for the person calling the room but not for somebody who is inside the room. The technician may also need to record multiple points in the audio pipeline and analyze the source of noise, a time consuming and difficult process. Even determining that the weird sound comes from a certain microphone may not be sufficient, as the cause may be due to broken hardware, e.g., microphone, cable, etc.; software or external factors (e.g., noise in the room that may only occur under certain conditions); etc. Such diagnosis can take substantial effort. Automatic failure alerts can provide a technician with sufficient direction on where to look, and may

even include detailed instructions on potential fixes.

3. Fixing broken rooms in response to a complaint is expensive - it disrupts meetings scheduled in the room during the time of repair, and prior to the repair, may cause lost meeting time where users attempt their own repairs. Automatically detecting failures early enables technicians to schedule repairs before end users are impacted, thus improving the utilization of meeting rooms, reducing the number of people required for maintenance, and the installation and/or repair time.

Example of room setup

For the purposes of discussion, assume a similar setup for all meeting rooms with video conferencing equipment. The setup includes at least one speaker and at least one microphone. Specific room 'types' or 'configurations' may have variations in this setup, for example, they may have a different number of microphones, microphone arrays and speakers. They may also have different geometries for the positioning of each of these elements. Additionally, each video conferencing unit is placed in a unique environment with specific acoustic characteristics that can impact its performance, such as the amount of reverberation in a room, external electrical interference from electromagnetic sources, acoustic interference from an HVAC system, etc. In different configurations, other equipment, e.g., a display such as a television, monitor, or projector; a camera; a light source; etc. may also be included in the room, in addition to the audio hardware. Fig. 1 shows the setup as described herein:

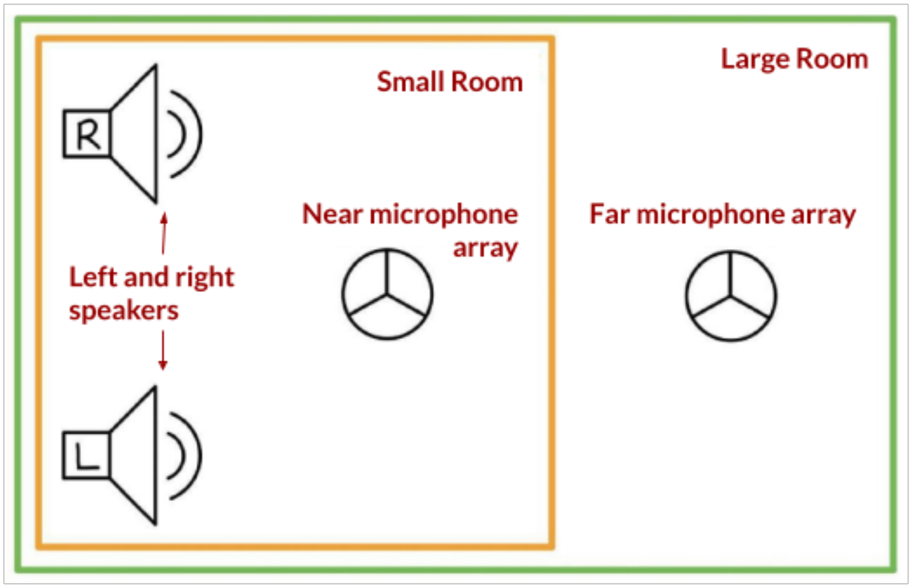

Figure 1: Example meeting room setup

As shown in Fig. 1, two example types of rooms are shown: (a) a “small” meeting room (shown with yellow border); and (b) a large meeting room (shown with green border). Further, assume that all meeting rooms are equipped with two speakers: a right one and a left one. “Small” has a single microphone array and “Large” has two arrays. Each microphone array includes three individual microphones. The video conferencing unit in “Small” and “Large” both have two ‘playback’ or output interfaces: one each for the left and right speaker. Each videoconference (VC) unit also has six ‘capture’ or input ports: Three for the microphones in the microphone array furthest from the speakers and three for the microphones in the nearest array. The capture ports in the “Small” configuration have the following arrangement: [mic_1, mic_2, mic_3, empty, empty, empty]. The capture ports in “Large” configuration have the following arrangement: [far_1, far_2, far_3, near_1, near_2, near_3].

<u>Common problems in meeting room audio peripherals</u>

With reference to the above setup, some of the common problems that can occur in meeting room audio peripherals include:

1. **Disconnected or broken microphones and speakers:** A common fault is a disconnected audio peripheral - a speaker or microphone is present in the room but doesn't function. This may be due to a fault in the peripheral itself (e.g., a broken microphone), a fault in the wiring (e.g., a loose connector), a fault within the video conferencing control unit itself (e.g., a broken connection port) or even a mistake in wiring (e.g., a connector plugged in an unexpected position).
2. **Obstructed microphone:** In general, it is expected that the audio levels from multiple microphones will have similar power (loudness) levels. A significant variation from the mean on a given microphone can be indicative of an occluded or obstructed microphone which requires manual intervention.
3. **Swapped capture channels:** At scale, it is important for VC system to maintain consistency in the setup and one source of variation is the mapping between microphones and the VC unit input ports. For example, in the example large room of figure 1, all microphones from the far array should be plugged into the first three ports and the microphones of the closer array into the last three ports. Mapping consistency is especially important for video conferencing units with stereo and multi-mic capture capabilities.

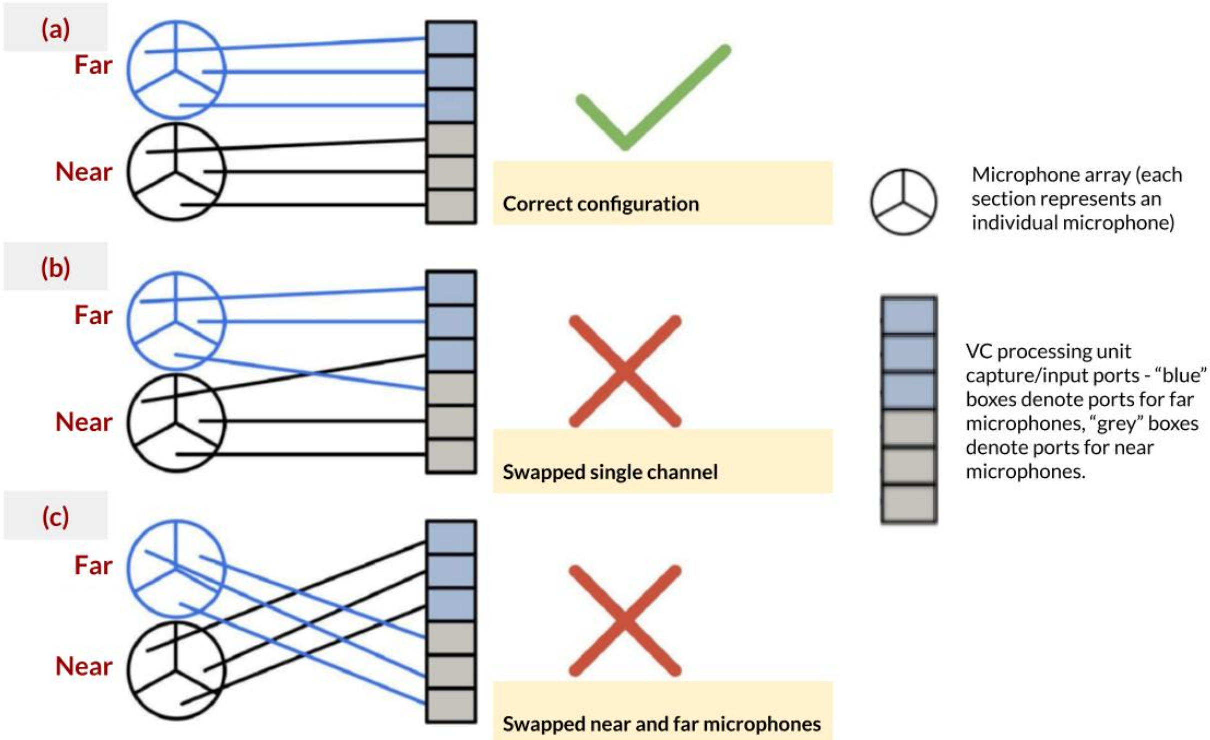

Figure 2: Example mappings of microphones to input ports

Fig. 2 illustrates various mappings of microphones from the near and far arrays to input ports. As seen in Fig. 2(a), microphones from the far array are connected to the first three inputs (shown in blue) and microphones from the near array are connected to the last three inputs (shown in grey), this is the desired configuration. In the example of Fig. 2(b), the microphones are miswired (one near and one far microphone is plugged into an incorrect port), while in the example of Fig. 2(c), the microphone arrays are swapped, with near microphones being plugged into the first three ports and the far microphones being plugged into the last three ports.

4. **Environmental interference:** Microphones are sensitive and may pick up undesirable sound from HVAC systems blowing air directly on to a microphone and this can be heard as a wind-like rumble. To avoid this problem, technicians can reposition a microphone or install wind covers. In addition, microphones can intermittently pick up rumble noises when air

flow patterns change, e.g., when HVAC systems go from cooling to heating in the winter, or the other way around in the summer. Further, some audio peripherals are also susceptible electromagnetic interference which may be picked up either directly, or via cable crosstalk if electrically noisy equipment is placed too close. While environmental interference isn't directly caused by a fault in an audio peripheral, it can degrade the quality of audio for a video conferencing system, requiring manual intervention from a technician and can thus be treated as another type of fault.

Fault detection techniques

Different types of faults can be detected by recording a reference recording (ideal sound) and comparing it with the sound which is actually picked up by each of the microphones (referred to herein as the microphone 'recordings'). Some examples of successful detection using this approach include:

1. Detect which of the speakers and microphones in a room are functioning. In the room examples shown above, the correct configuration is two working speakers and either three or six working microphones (depending on whether it is a small room or a large room).
2. Determine the number of microphone arrays in a room
3. Measure the average 'loopback' delay between signal generation to signal received in the room test software. The loopback delay is the sum of: transfer delay of audio data from software to the audio hardware and speaker, time of flight for sound waves in the room from speaker to mics, transfer of audio data from microphone audio hardware back to software and delays in the operating system.
4. Detect if the microphone arrays have been swapped, e.g., as shown in Fig. 2(b) and 2(c).

5. Detect if the microphones have been miswired

These are explained in detail below.

Detect working microphones and speakers

To detect the microphones and speakers that are working, the speakers are used to (a) play a loud sound through the left speaker; (b) wait for a delay time until the system has achieved a steady state; and (c) play a loud sound through the right speaker. During each of three phases, all the audio inputs are recorded. If a signal above a certain loudness threshold is detected in any of the recordings of the audio inputs during the first phase, it is inferred that the left speaker is functional. The same holds during the third phase with the right speaker. For any input that detects a signal during any of the phases, it is inferred that a microphone is functional. Detecting which microphones and speakers work is valuable information since all other tests require at least one working speaker and one working microphone. If it is determined that no microphone and/or no speaker works, then the test session is aborted and an alarm is raised.

Determine the number of arrays

Based on the results of the previous test, it can be concluded that there is a single microphone array if there are exactly three working microphones and two arrays if there are six working microphones. Any other detected number of microphones is invalid and can be used to raise an alert. Additionally, it is determined if the number of detected arrays matches the expected number. In the examples of this disclosure, a small room should be detected as having one connected array and a large room as having two connected arrays. Further, a working microphone can detect all working speakers and a working speaker is detected by all working microphones.

Detect swapped arrays based on time-of-flight

Testing for swapped arrays in a small room requires a different process to that of a large room. In a small room, the first three input ports are to be connected to the microphone array and the last three ones should be empty. Since the connections between input ports and working microphones are known, this is straightforward. As shown in Fig. 1, in the large room, the first three input ports are connected to the far array and the last three ones to the near array. For simplicity, it can be assumed that a certain distance (e.g., 5m) exists between both the left speaker and the right speaker and the near array, and a larger distance (e.g., 10m) exists between the speakers and the far array. Fig. 3 shows this configuration.

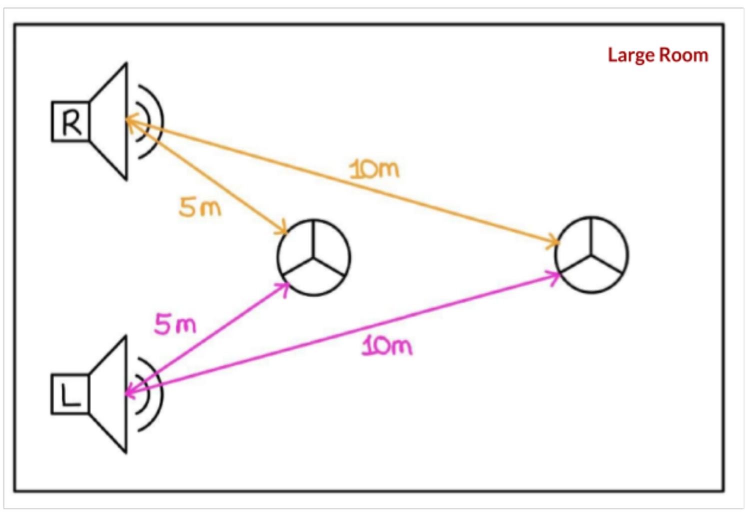

Figure 3: Detecting swapped microphone arrays based on time-of-flight

The speed of sound can then be used to calculate the expected time of flight from the speaker to the microphones as follows: assume speed of sound: 343 m/s (nominal air pressure, humidity and temperature). Then, the time of flight for the first array is 5m / 343m/s ≈ 11.5ms and the time of flight for the second array: 10m / 343m/s ≈ 23ms. The delta between the two arrays, in this example, is therefore 23ms - 11.5ms = 11.5ms.

There are some factors which make the measured time of flight values less precise, such as:

- the microphones inside the array point into different directions
- the arrays might be (partially) obstructed
- the time of flight depends on the ambient temperature
- measuring the onset of a sound wave to a microphone incurs a tradeoff between accuracy in time and rejection of false positives due to random ambient noise (better time resolution means less resiliency to transient noises)
- the ‘loopback delay’ measurement also includes speaker and microphone processing times, in addition to the time of flight, with additional jitter introduced by different components.
- Further, the delays for one microphone might be a bit longer or shorter than others.

Even with such inaccuracies, it can be determined that the arrays are wired in the correct order if the average delay of the first three microphones is higher than the average delay of the last three microphones.

Detect miswired microphones

This test requires differentiating between small rooms and large rooms. In the correct configuration, small rooms have three working microphones that are connected to the first three input ports. The knowledge of the speakers and microphones that are working (as detected above) can be used to test if the microphones in the small room are miswired. Detecting miswired microphones in large rooms requires a different approach. The average loopback delay of the first and second array (as measured above) is compared with the loopback delay of each individual microphone. If the delay of an individual microphone is not consistent with the delay of its expected array (based on input port), the microphone can be inferred to be miswired. For example, a microphone which belongs to the near array should have a delay which is closer to the average delay of the near array than to the one of the far array. The same holds for a microphone of the far array and the average delay of the array.

Interference detection

Interference detection is based on directly analyzing the audio signal from a room, no reference signal is needed. Reliable results can be obtained when there are no loud external room sounds, e.g. a person using the room, that interferes with the test. Even a microphone in an empty, seemingly silent, room may pick up random noise which is a mixture of noise from different sources: electromagnetic (e.g., cross talk and RF), mechanical coupling (e.g., vibrations from heating, ventilation, and air conditioning (HVAC) systems) and thermal effects etc.

This background noise is amplified and passes through a series of filters (high pass, low pass and noise suppression filters) which attempt to remove it. While such filters are effective in removing the interference, the filters may introduce new sound artefacts and degrade audio quality in the process. Figuring out that degraded audio quality in a room has been caused by HVAC noise, for example, requires a significant time investment and experimentation by technicians. Therefore, it is important to automatically detect rooms with abnormal and/or loud background noise level.

One method to automatically detect rooms with abnormally high background noise is to use the root mean square (RMS) of each microphone's signal. The RMS level provides a good first indication of how loud the background noise in a room is. Given n audio sample values the RMS is defined as follows:

$$\text{values: } \{x_1, x_2, x_3, ..., x_n\}$$

$$\text{rms} = \sqrt{1/_n * (x_1^2 + x_2^2 + x_3^2 + ... + x_n^2)}$$

The RMS signal level provides an indicator of room volume level; however, it does not always represent how "silent" a room sounds to a human.

Classification via RMS & frequency strength distribution

It is observed that ‘silent’ sounding rooms follow a characteristic distribution of energy over frequency. A room with a higher RMS which follows the frequency strength distribution of a ‘silent room’ usually sounds much better to humans than a room with has a lower RMS but does not follow the nominal frequency strength distribution. Accordingly, rooms can be classified into one of four categories, each of which is indicative of the room’s characteristics.

1. Below RMS threshold & follows frequency strength distribution. A room in this category is ideal.
2. Below RMS threshold & does not follow frequency strength distribution. This is an unlikely category in practical settings.
3. Above RMS threshold & follows frequency strength distribution. A room in this category is not ideal but most likely sounds good enough.
4. Above RMS threshold & does not follow frequency strength distribution. A room in this category does not sound good. There is probably some sort of disturbance or error, e.g., a bug in the recording, HVAC, person in the room, building work nearby, broken hardware, or another reason.

For the frequency strength analysis, the frequencies are grouped together in buckets of exponentially increasing size, e.g., bucket 1 -.0-500Hz, bucket 2 - 500-1000Hz, bucket 3 - 1000-

2000Hz, bucket 4 - 2000-4000Hz, bucket 5 - .4000-8000Hz, and so on. This distribution approximates the way human hearing works (approximately following a Mel scale) and thus creates buckets of roughly the same (perceived) size. A Fourier Transform (FFT or fast Fourier transform) can be used to obtain the energy content in each frequency bucket in the recording. The results can be plotted, with frequency buckets on the x-axis and the average strength of all frequencies that belong to each bucket on the y-axis. High values can then be excluded using a threshold (which can be different for different buckets), with values below the threshold being considered acceptable.

Considering that a single microphone may malfunction, or that the recording might not follow the expected distribution because a single sample value may be above the threshold of its bucket, a microphone array is identified as noisy only if at least half of the microphones of the array do not follow the expected frequency strength distribution. A room which includes at least one noisy array is considered noisy.

As described the analysis of the recordings classifies the rooms noise level into four categories, however, by extension, additional tests can be carried out to detect other attributes such as: whether a single microphone in the room is broken? (independent of the noise level in the room); does the room have a periodically recurring sound?; Does the frequency spectrum of the room differ from the one of a silent room and why, e.g., due to HVAC, a single loud sound, human voice, or some other reason?; Did the room fail multiple consecutive tests for the noise floor?

CONCLUSION

This disclosure describes audio selftest techniques that can automatically detect faults in video conference hardware and enable preemptive repair of the detected faults to ensure that meeting rooms maintain a minimum set of quality standards. The audio selftest includes the playing of sounds through in-room speakers and detecting the sounds via in-room microphone arrays. Common problems such as disconnected or broken speakers/ microphones, swapped microphones or microphone arrays, miswired microphones, etc. can be diagnosed. Further, interference in the room can be detected and rooms can be classified as suitable or unsuitable for meetings based on ambient noise detected in a silent room by classifying the detected signal. Such automated diagnosis can enable frequent testing of meeting room hardware, provide early detection of malfunctioning equipment, and enable faster repair.