XEKERA SYSTEMS

Solutions Simplified

SUITE OF TESTS FOR AV (AUDIO-VIDEO) SYNCHRONIZATION

REAL TIME AUDIOVISUAL EVALUATION SUITE

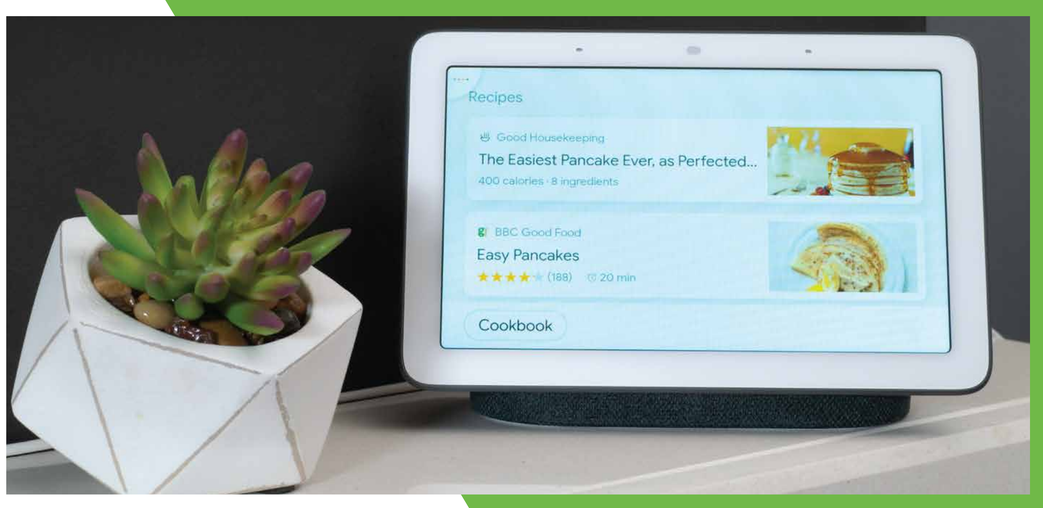
INTRODUCTION

One of our major customers at Xekera Inc. approached us with a need for devising a system test solution for evaluating the audio-visual performance of content playback from the cloud on a number of different devices. Xekera did not have tools developed in this domain, but given our expertise and experience in system testing for consumer electronics, we accepted the challenge to architect and implement the solution per the customer's need.

This case study goes over the challenges, the solution, and the Key Performance Indicators (KPIs) that we focused on for this solution. We will keep the scope of this study at a high level for quick and easy comprehension, and are willing to discuss the specifics of the implementation with any interested party as needed.

THE STORY OF THE CUSTOMER

Our customer’s offerings and devices cover the lion's share of their market, but their internal test engineering focus is to create services and solutions which can scale on as many products as possible.



Such an approach requires their testing teams to be focused on unit and integration testing, which are much more scalable for their continuous integration and continuous development (CI/CD) methodology. Xekera needed to come up with a system test solution which uses the same system boundaries as the real users and can be integrated with customers' existing CI/CD framework as well. The system test solution is also required to be scalable to future and legacy products.

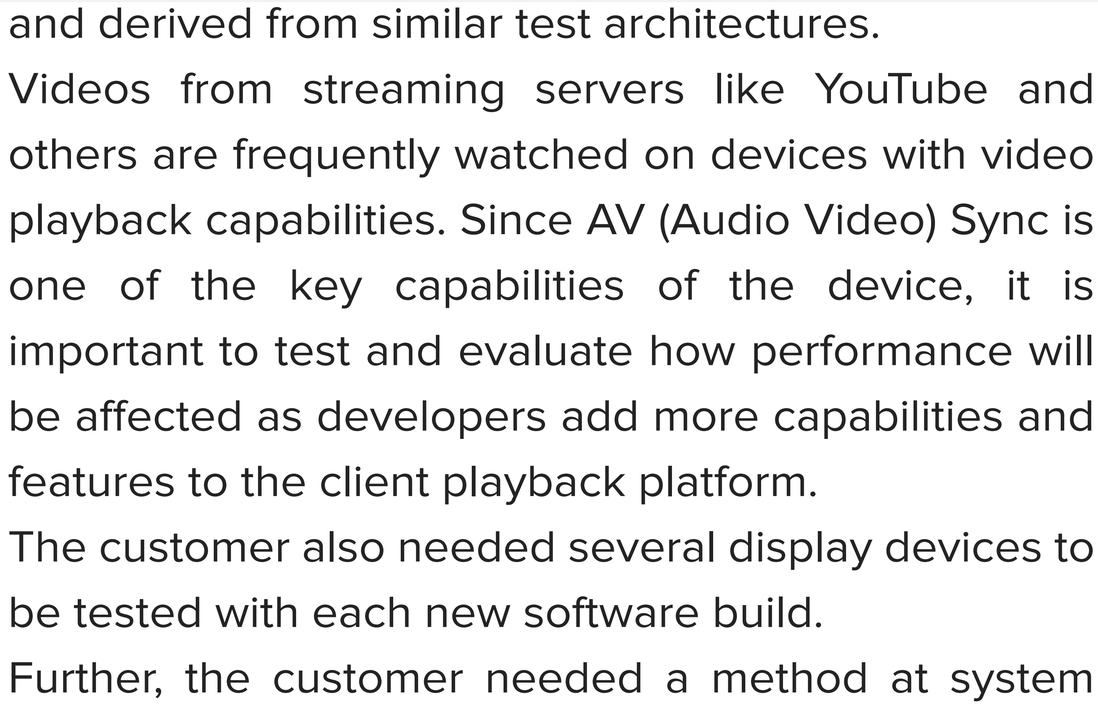
THE CHALLENGE

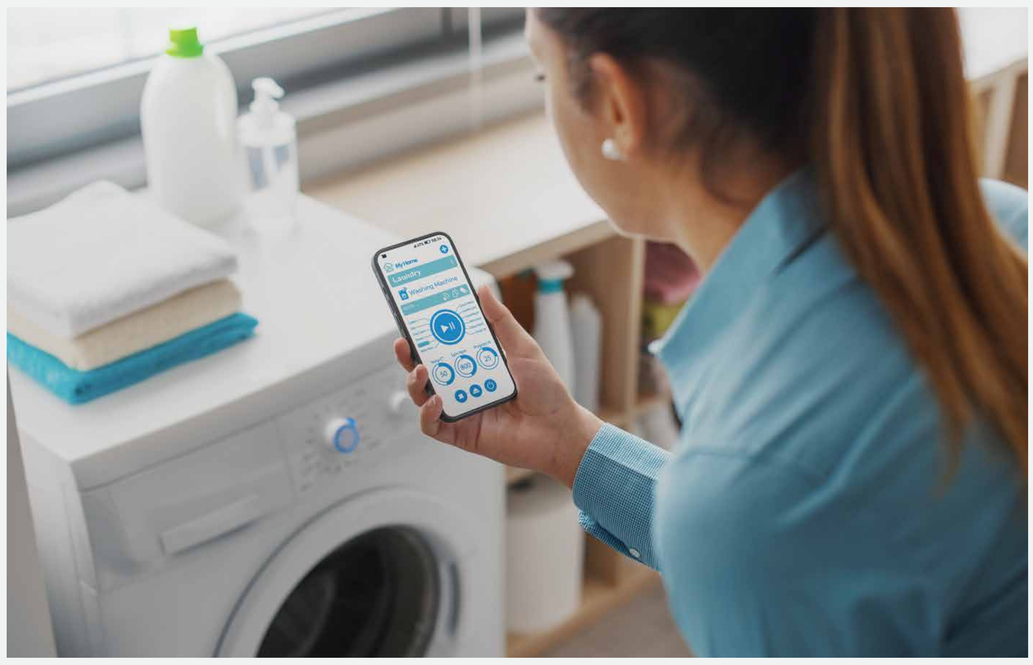
In a modern video streaming service, audio and video packets are sent to the consumer-end playback devices separately, and the client application running on the end device is responsible for stitching the packets together in a way that users experience the playback without any issues of audio-video synchronization (lipsync). This principle is one of the bases for Xekera’s solution development in this area, and this study will use this basic KPI to explain the solution, and the architecture of the solution. We will then also illustrate how other KPIs can be measured and derived from similar test architectures.

Videos from streaming servers like YouTube and others are frequently watched on devices with video playback capabilities. Since AV (Audio Video) Sync is one of the key capabilities of the device, it is important to test and evaluate how performance will be affected as developers add more capabilities and features to the client playback platform.

The customer also needed several display devices to be tested with each new software build.

Further, the customer needed a method at system level to precisely gauge how consumers perceive the offset between audio and video outputs.





WHY CUSTOMER CHOOSE XEKERA SYSTEMS

Xekera Inc. has proven systems-level solution development capabilities. We have a history working with several of our customers developing hardware, software, and firmware, as well as performing calibrations, all at full-stack level. Xekera has developed products from definition to production, and our customers have full exposure to our capabilities and facilities. Owing to some prior work that we did for this customer showing our capabilities, they gave Xekera the first opportunity to study and propose a solution. Our proposal was received very well, being much more comprehensive and extensible in comparison to other solutions pitched to our customer.

HOW XEKERA SYSTEMS RESPONDED: THE SOLUTION

Any device with a display is within the purview of this test solution. The test technique aims to evaluate the streaming user experience, where the device’s display and speaker are crucial factors. In order to evaluate the streaming experience, the device's audio and video output need to be recorded for analysis.

The aim is to playback an objectively created test stimulus that contains audio and video markers that can be captured for offline evaluation as the test stimulus is reproduced or played back on the target devices. As a first step, we designed a test stimulus in a way that any desynchronization (desync) between audio and video in the signal recorded from the playback device can be measured. This test stimulus has to be designed with consideration of the refresh rate of the playback device.

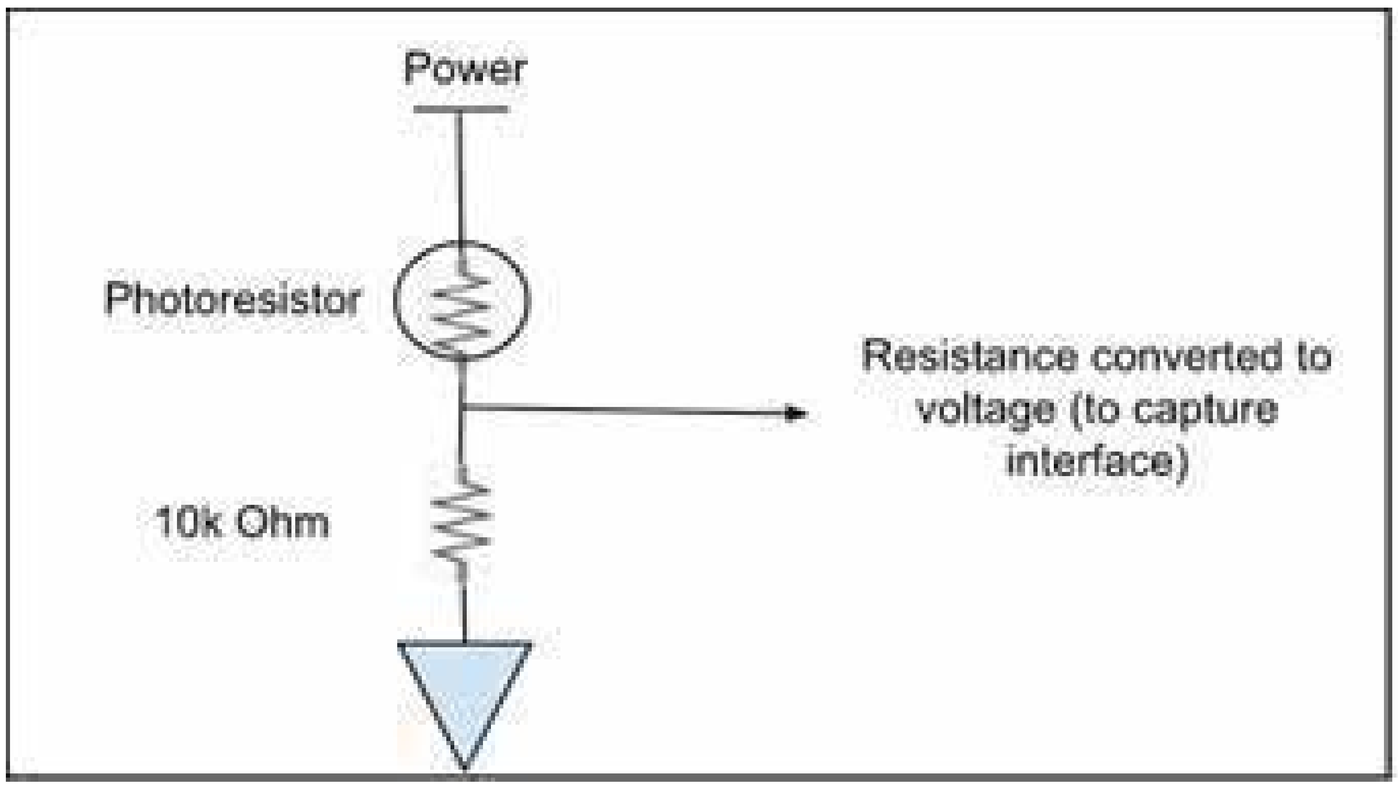
To address the test stimulus requirements, we produced a 60 fps test stimulus with 3 frames of white and 57 frames of black. A 1 kHz test tone is placed in in sync with white frames (ensured as part of test stimulus creation).

Fig 1: White and Black blinking frames video

Note that the diagram above shows a frame with split color blocks (2 coded blocks per frame). This facilitates pick up of the differential video signal as well as other experimental considerations. Either section of the frame above can be used to record the sensor output, depending on the desired duty cycle of the waveform.

The second step was to capture the playback once the test stimulus is played on device. We chose to record device audio in the electrical domain, or "conductive audio," to assure the best signal-to-noise ratio and scalability of the system without the need for audio isolation. In order to enable conductive audio capture, the playback device was modified by rerouting the speaker signals to an audio interface board of our own design. This audio interface board designed by Xekera was essential to condition the signal from speaker-level outputs to line-level inputs, and the differential to single-ended conversion.

For video capture from the playback device screen, the light energy must be transformed into electrical energy so that evaluation can be performed on recorded audio & video signals. We used a photoresistor placed on the playback device screen as a sensor to create varying electrical signals according to the screen brightness (depending on white or black frames). A USB sound interface is used to capture both audio output from a playback device and voltage signals from a screen sensor, as they are now both in the electrical domain. The following circuit is used to convert display frames to voltages. The sensor board containing the photo resistor was designed by Xekera for this specific solution, as shown below.



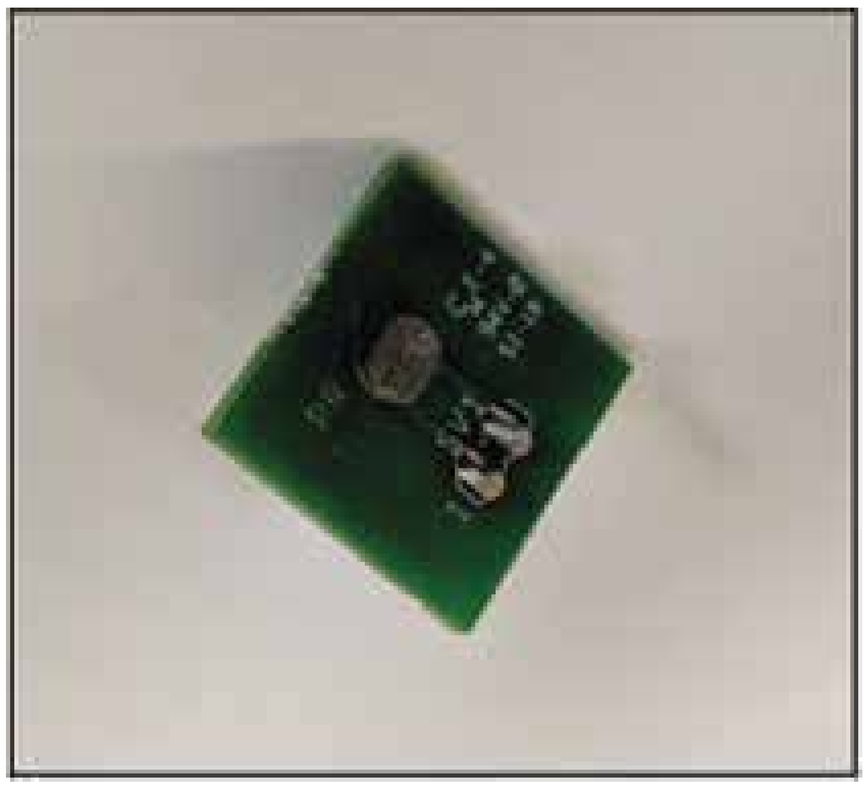


Fig 2: Photo sensor to measure video signals

A complete system-level test setup is shown in the diagram below.

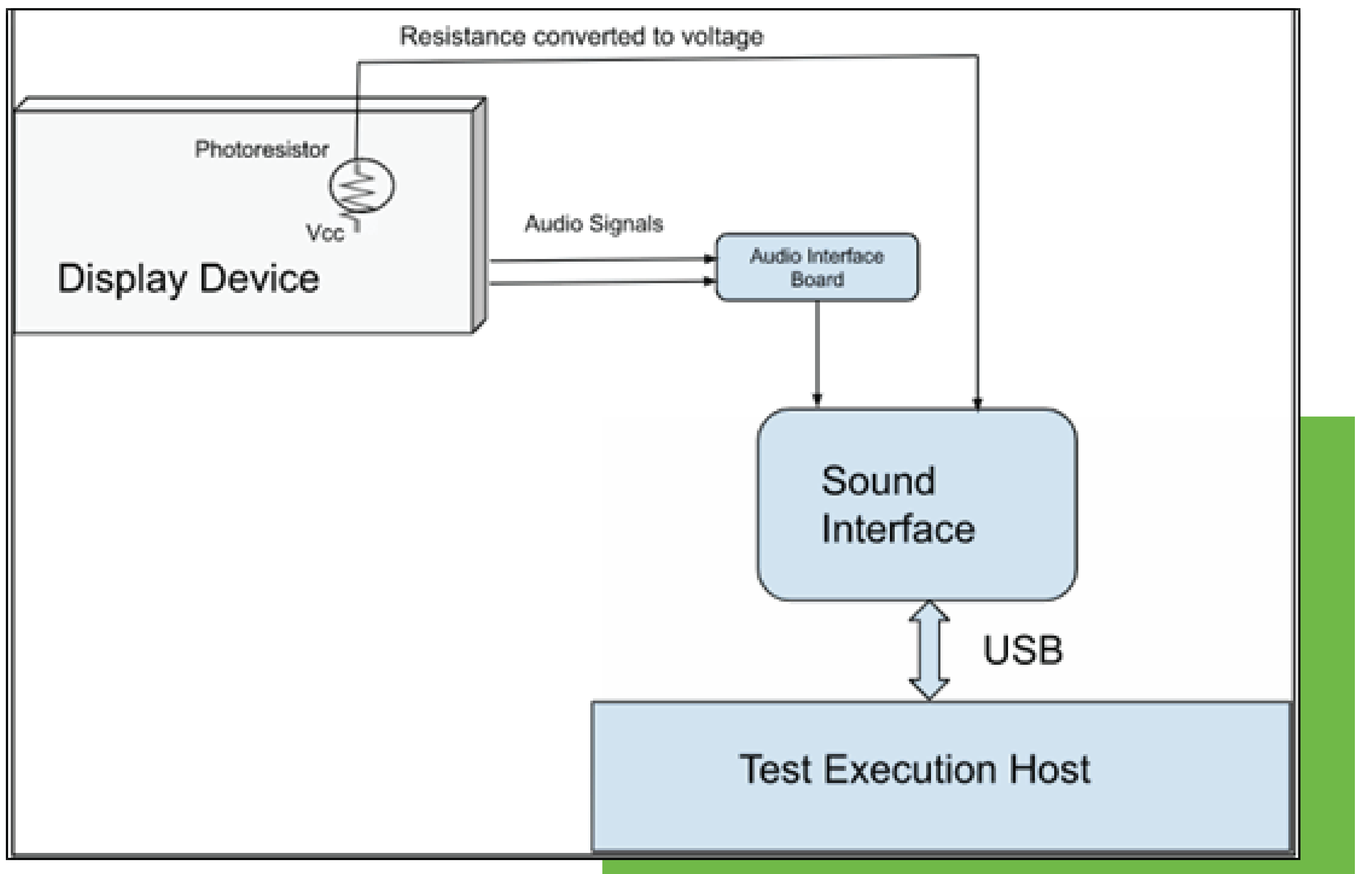


Fig 3: Testbed Setup block diagram

This testbed provided the solution for test stimulus playback & capture with fidelity. This can then be analyzed by software algorithms.

One of the challenges during the development of this solution was to calibrate the playback device audio gains and screen brightness to achieve a desired signal-to-noise ratio (SNR) for analysis by software. Xekera utilized its system expertise to develop tools that calibrated the device gains by iteratively playing back & recording the audio/video signals until the desired SNR was achieved.

Although details of the calibration algorithm are not in the scope of this discussion, we show below the output of the calibration process for ease of understanding.

```
    ***Recording Started***
Recording WAVE '/tmp/AVsync_mc_output.wav' : Signed 16 bit Little Endian, Rate 48000 Hz, Channels 8
   ***Recording STOPPED***

 ****AV analysis ******

CH0 Audio peak mean is -9.83334dB,it is within threshold of (-8dB to -9dB)
+-----------+-------------+----------+----------+----------+
|  Channel  | total peaks |   mean   |   max    |   min    |
+-----------+-------------+----------+----------+----------+
| Channel # |      9      | -9.83334 | -9.80288 | -9.85466 |
+-----------+-------------+----------+----------+----------+

CH1 Video peak mean is -2.89313dB,it is within threshold of (-4dB to -6dB)
+-----------+-------------+----------+----------+----------+
|  Channel  | total peaks |   mean   |   max    |   min    |
+-----------+-------------+----------+----------+----------+
| Channel # |      7      | -2.89313 | -2.88579 | -2.89615 |
+-----------+-------------+----------+----------+----------+
```

Fig 4: Calibration tool output

Once the setup is calibrated, the final step is finding the offset between the audio and video recorded signals. It is accomplished by utilizing algorithms developed in-house at Xekera. The analytical algorithm used for signal processing is discussed further in the section below.

This algorithm analyzed the recorded signals, their properties and amplitude, and displayed them to user-readable results in an automated manner with a tool developed in Python. This allowed our customer to continuously evaluate the audio-video desynchronization produced by the devices in different use cases (playback, rewind, fast forward, pause).

SIGNAL PROCESSING

The offset between the audio and video (desync) is determined by measuring the offset between the start of a white frame and the start of the corresponding 1 kHz tone. All the calculations are performed based on the samples and according to the audio interface sampling rate.

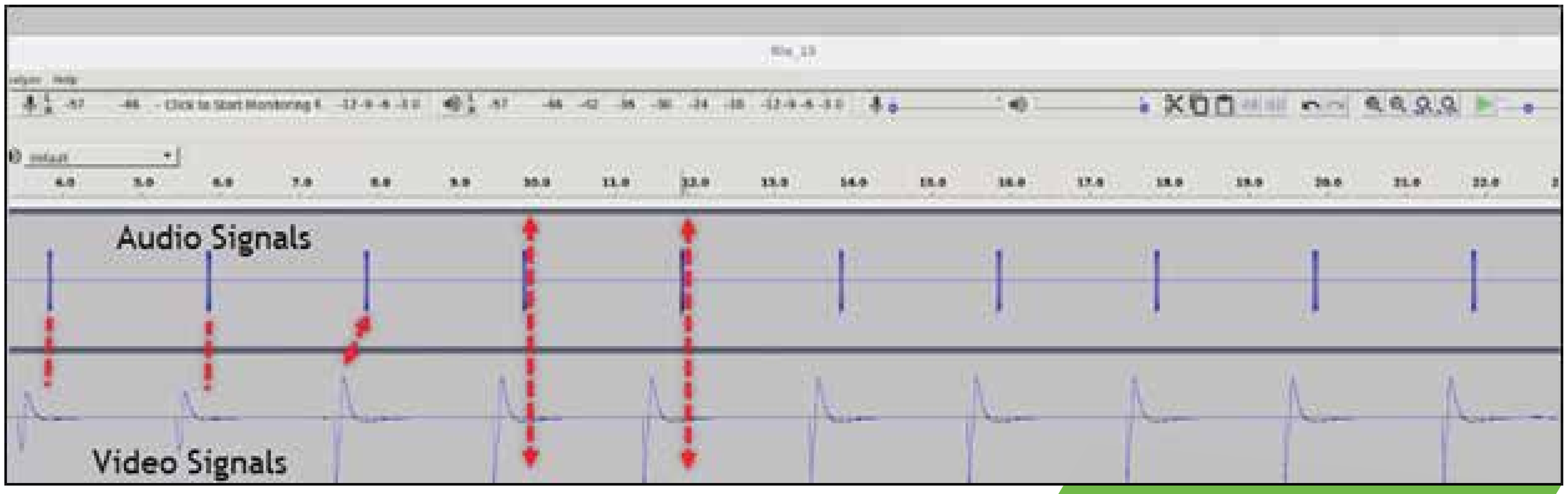


Fig 5: Desynced audio and video signals

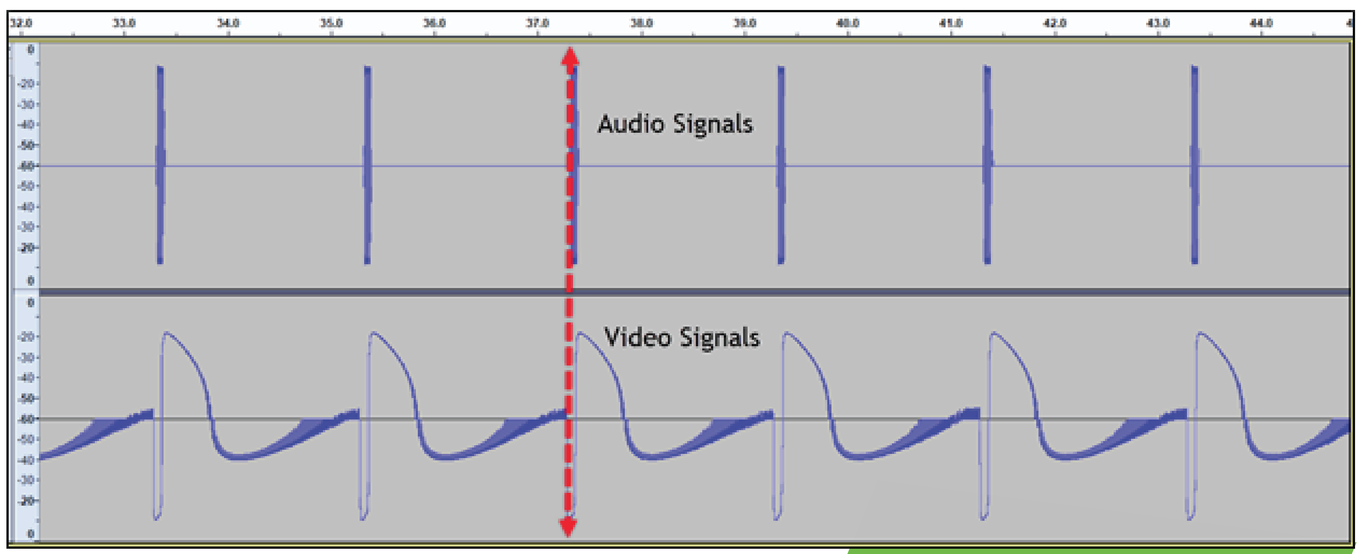


Fig 6: S ynced audio and video signals

The solution developed by Xekera is extensible and was further enhanced by adding more coded sections and photo sensors on the same display screen, enabling the measurement of several other KPIs like freezing (missing or static frames) and smoothness (uniformity of frame distribution), two important metrics for streaming quality. As an example, we enabled the monitoring of the following KPIs (as well as others)

- Smoothness of the Output frame distribution - i.e. frame frequency
- Freezing i.e. the number of consecutively missed frames.

The additional coded sections in the test stimulus change color at different rates, varying from 30 frames (0.5 s) white and 30 frames (0.5 s) black to changing white and black alternating frames (16.66ms). This allows us to capture the change in frame color at different rates and thus determine the measured screen refresh rate against the expected refresh rate. Smoothness of video playback can be easily calculated using algorithms on recorded waveforms from each sensor. Similarly, by observing the variations in duty cycle of the recorded waveforms in different channels, we can determine how long an instance of freezing lasts. The diagram below shows the test setup using multiple sensors/coded sections in test stimulus, as well an example captured waveform from multiple sensors on the same playback device.

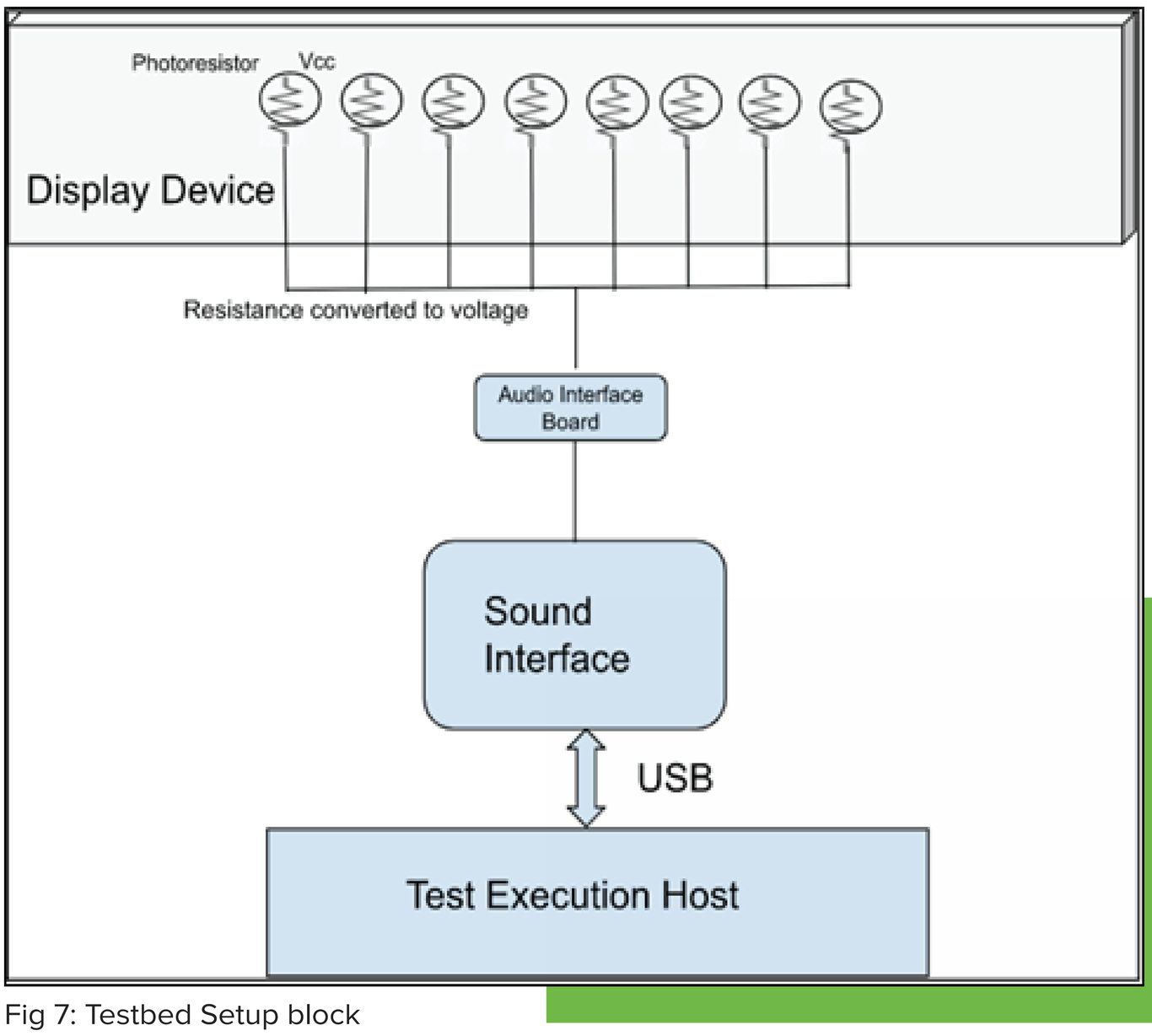


Fig 7: Testbed Setup block diagram | Multi- sensor

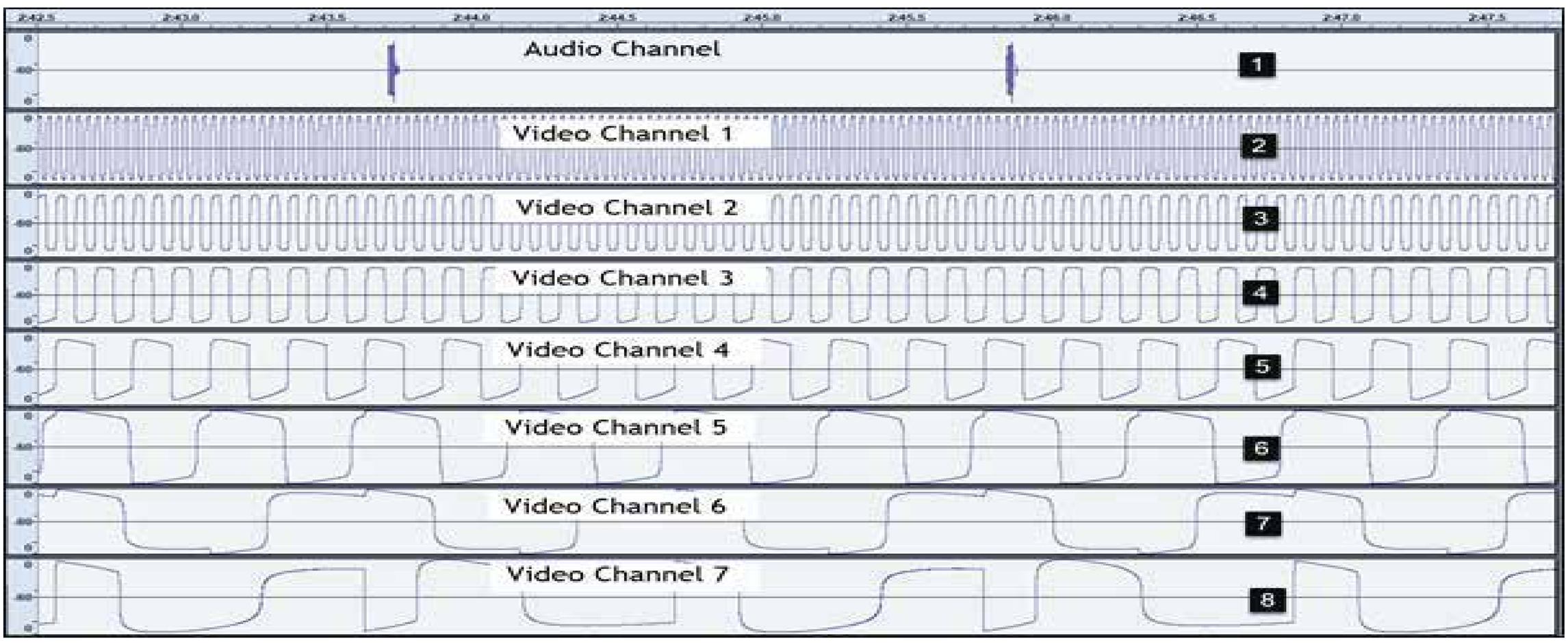


Fig 8: Recorded wav file .

Xekera developed software algorithms to process the recorded waveforms using our in-house signal processing expertise. The KPIs mentioned above illustrate the concept, and several other KPIs are calculated using the same recorded waveform. This methodology allows us to perform the test run once and determine multitude KPIs from the same captured waveforms, which is very valuable from a test complexity perspective. The design and logic of the signal processing algorithms is out of the scope of this discussion but can be discussed in detail with our design team as needed.

THE RESULTS

Xekera's test solution demonstrated the capability to gauge the following KPIs by integrating various light sensors:

- Smoothness: Using output frame distribution.
- Freezing: Using the total number of consecutively missed frames
- AV Sync: By adding various frequency tones to the audio channel to measure AV Sync with greater accuracy.

The work by Xekera that is described here enabled an automated suite of AV tests to be executed for every new build of the device by the customer. This test methodology is scalable and can be performed on any display device with video playback capability. Currently, the customer has this solution deployed in their labs for legacy, as well as to-be-released devices. This automated process aids the customer in identifying any problems or validating a new device or updated firmware before launching the device or releasing the firmware.