

About the Project

Audio Prototype

The customer provides end-to-end video streaming solutions for encoding, recording, managing, publishing and distributing video content. Their video streaming solutions power a wide range of applications that meet their customer requirements.

As interest in VR-themed solutions was increasing, solutions that allow recording and streaming video in 360-degree format began to appear. However, in addition to entertainment, this format has the potential to be used in business solutions. Thus, the customer had the idea of applying 360-degree format to interactive videoconferences.

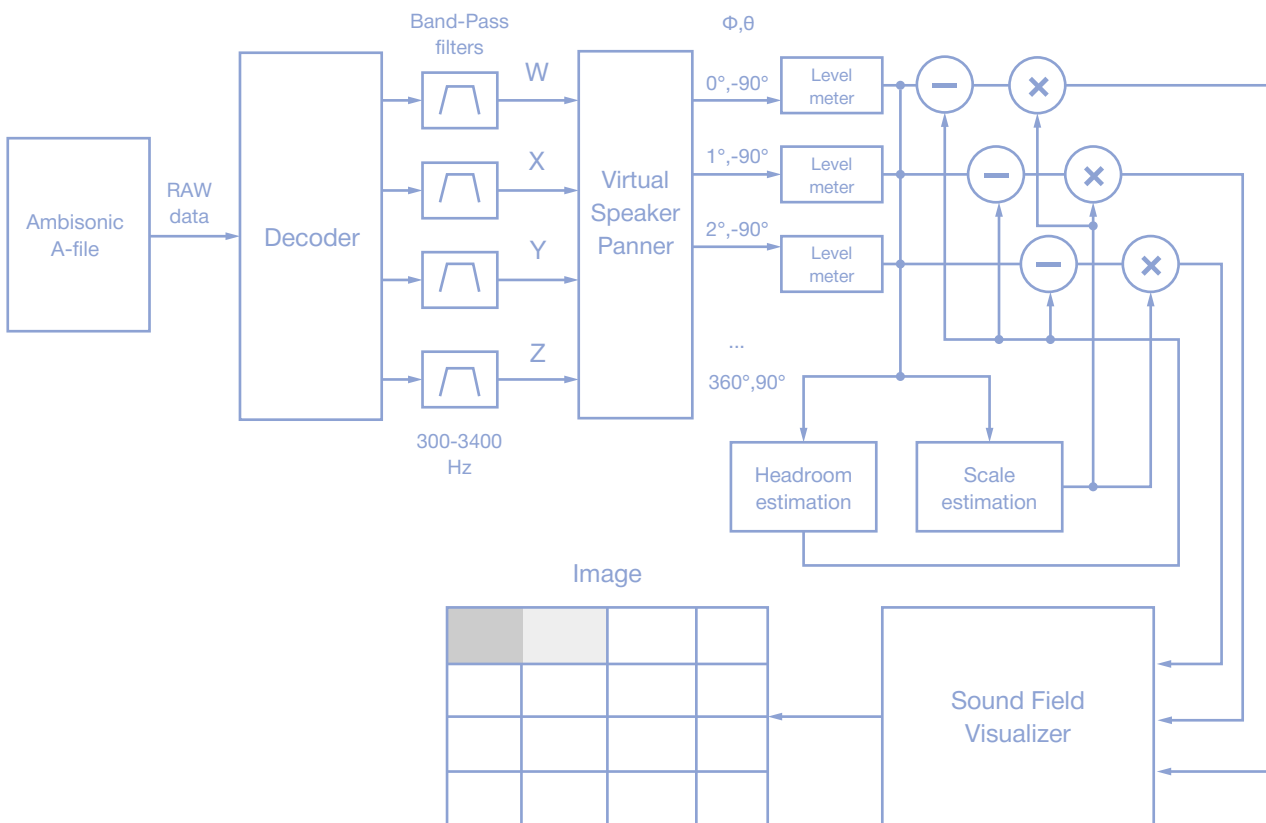


Challenge

The main idea is that participants could be placed randomly in front of the video camera. This would significantly improve the usability of the entire solution. Thus, there is no need to build complex schemes of multiple cameras and additional equipment to place all the participants.

In addition to the particular 360-degree camera, our customer needed to implement a component that allows the user to detect the position of the speaker in live-mode and to transmit this data to the other side for which the playback is being performed. This is necessary so the participants on the other side can see the person who is currently speaking.

Detector operation scheme



Solution

To implement the component for detecting the speaker, together with a 360-degree video camera, we decided to use a special microphone that allows the recording of spatial audio in Ambisonics A-format



Then we needed to investigate and build a software solution (algorithm) for processing Ambisonics A-audio, which allows us to detect and calculate the direction vector to the loudest audio source. For that, we used **the algorithms FFT, Convolution, AGC, HRTF**, as well as a number of algorithms for signal processing of **the OpenCV library**. The main idea of this approach was to build a sound field map in polar coordinates.

Further, using algorithms for digital image processing (Threshold, Erode, Dilate, contour detection), the coordinates (the direction vector) of the loudest sound source are analyzed and calculated.

Features

The app allows the user to:



analyze Ambisonics A-audio from various sources (sound card, audio file, HLS stream)



implement metadata containing direction data into H.264 video stream using Wowza WMS



perform vector detection and calculation in live-mode



generate various debugging information, and visualize the 360-degree map of sound field levels.

Technologies Used

Digital Signal Processing
(HRTF, FFT, Convolution, AGC)

PortAudio

ZeroMQ

FFMPEG

OpenCV

Wowza WMS

 **FFMPEG**

PortAudio

 **ZeroMQ**

Digital Signal Processing


OpenCV

 **WOWZA**
media systems