IDEAS
Improving digital encoded audio skills

MODULE C

COMMUNICATION THEORY

GIANLUCA PATRITO

LATENCY MANAGEMENT

Latency, or temporal delay, is a critical element to consider in audio and video streams, aiming to avoid situations where sound precedes or lags behind visual action.

Latency compensation techniques include optimizing transmission protocols, employing advanced compression algorithms, and implementing intelligent buffering. The primary goal is to minimize the temporal gap between recording and playback.

Addressing latency in audio and video streams and ensuring accurate synchronization are crucial aspects for delivering a high-quality multimedia experience, especially in consumption contexts where timing and alignment are essential.

AUDIO LATENCY

More specifically, it is the time between when an incoming signal is fed into a recording device, a digital audio workstation or a streaming software program, and when the outgoing signal is emitted by loudspeakers or played through headphones.
This delay is usually measured in samples per millisecond.

Latency can also be defined as the time it takes an audio signal to:

- Get fed into a preamplifier;
- Access the audio interface;
- Be subject to analogue-to-digital conversion;
- Be acquired by a streaming software program or DAW;
- Get read and processed by the streaming software program or DAW;
- Be subject to digital-to-analogue conversion;
- Come out from loudspeakers or headphones.

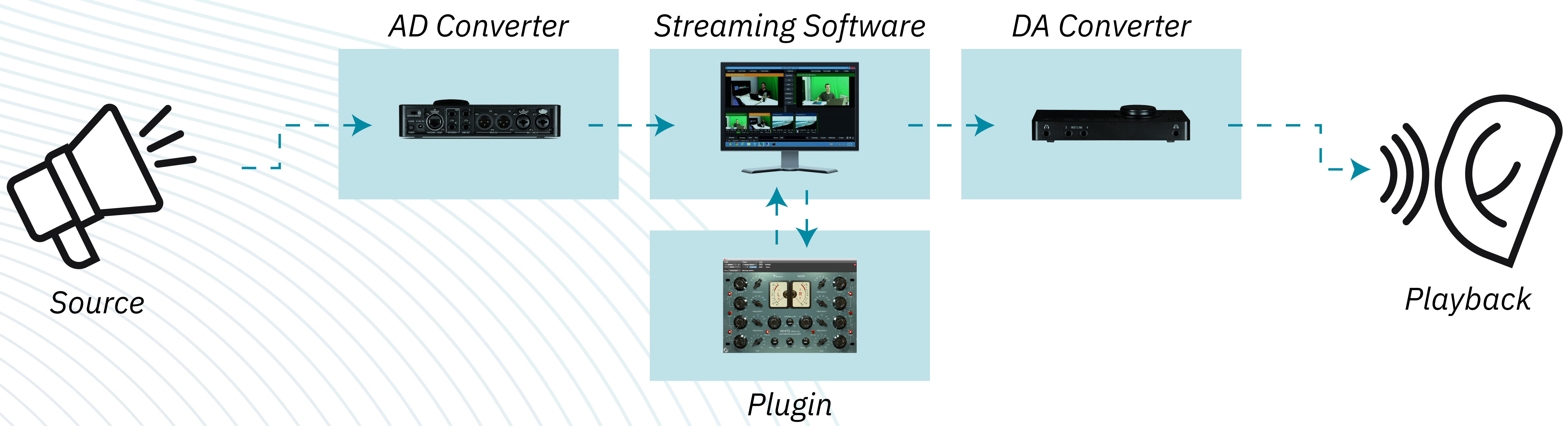
SIGNAL ACQUISITION AND PLAYBACK LATENCY

During the signal acquisition phase, a signal travels through an AD audio interface. From there, it is fed into and processed by a streaming program or a DAW, and then it is fed into and reproduced by a DA audio interface.

Latency is generated by the analogue-to-digital and digital-to-analogue conversion phases, by the sample rate and buffering process, and by the transition from hardware to software.

It is worth noting that opening and running other programs and apps requires additional CPU and RAM usage. This may affect the time it takes a signal input to reach its destination.

LATENCY IN AUDIO ACQUISITION



Audio latency during the signal acquisition phase and the playback phase can be caused by:

- Audio interface settings and configuration;
- Computer hardware specifications (suboptimal RAM and processor performance);
- Streaming software or DAW settings;
- Signal processing via plug-ins.

ENCODING LATENCY

An audio codec (encoder or decoder) is a computer program or device which is used to encode (write) or decode (read, playback) an audio data stream.

Audio codecs are used for audio file format conversion and are generally what many editing programs – such as audio and video editing software, audio converter tools, and streaming software – rely on.

Codecs compress data, thus reducing the amount of data points in an audio stream.

This compression can either be lossy (i.e. with loss of information) or lossless (i.e. with no loss of information).

For example, in audio streaming, one of the most popular encoding operations is from WAV to ACC or MP3.

LATENCY COMPENSATION IN AUDIO SIGNAL ACQUISITION

The audio buffer is the region of memory used to compensate differences in transfer speed and data transmission speed. Buffer size is measured in samples.

Low buffer values – generally seen in signal acquisition – correspond to swift transfer and transmission times. Latency is minimal in such cases.
Higher values, which emerge during processing operations involving multiple plug-ins, usually indicate longer transfer and transmission times; latency times here are substantial.

During acquisition, latency can be compensated by optimising buffer settings.
In USB and Thunderbolt audio interfaces, latency is measured in samples and is also referred to as “audio buffering”. On the other hand, in audio interfaces which use Ethernet for data transfer, it is measured in milliseconds.

VIDEO LATENCY

Video latency refers to the period of time between when an input signal reaches a recording device, a video control room, a vision or a video mixer, streaming software or hardware, and when it reaches an output device which can reproduce this outgoing signal – such as a computer, a streaming device, a Led wall, a monitor, etc.
This delay is usually measured in frames per millisecond.

The transition of the video signal to hardware is what causes latency. Such pieces of hardware could be:

- Vision mixers
- Video mixers
- Protocol converters

VIDEO LATENCY

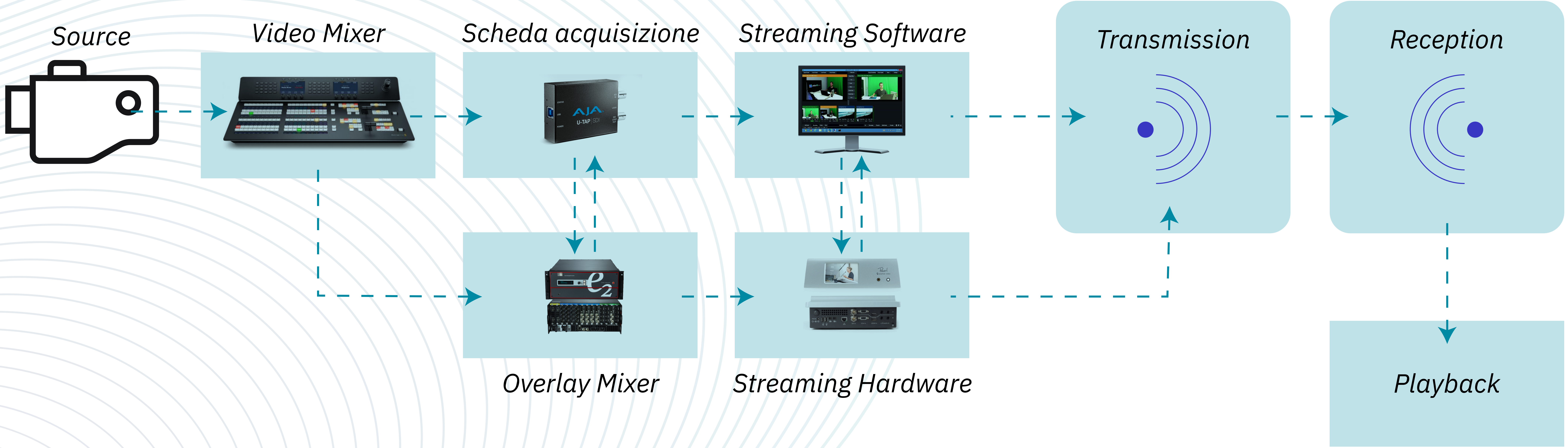
When processing a 1920x1080 video with 25 frames per second, 1920x1080 pixels will be generated 25 times a second.
Given that a frame is equal to 1/25 of a second, there will be a 40-millisecond delay in the video signal compared to the audio signal.

Before being streamed or reproduced on a playback device, a video must be processed and edited using the equipment we listed earlier. Higher hardware performances equate to lower processing times and hence lower latency.

Hardware quality plays an important role in determining total latency.

Digital video latency is the period of time delay between input and output. This delay becomes visible when we compare the frame at the source with the frame being reproduced.

LATENCY IN VIDEO ACQUISITION



The video signal is captured in high definition and will be compressed to reduce bandwidth usage, introducing additional delay to the signal processing.

AUDIO TO VIDEO SYNCHRONISATION

Despite both audio and video processes incurring latency, the element that will consistently be more delayed between the two is the video due to its higher data intensity, requiring more time for processing and encoding.

In order to synchronize the two sources, it will be necessary to apply a delay to the audio signal.

To determine the standard latency, considering a single element that introduces latency, the formula to be applied is as follows:

1000 ms / FPS = Latency (ms)

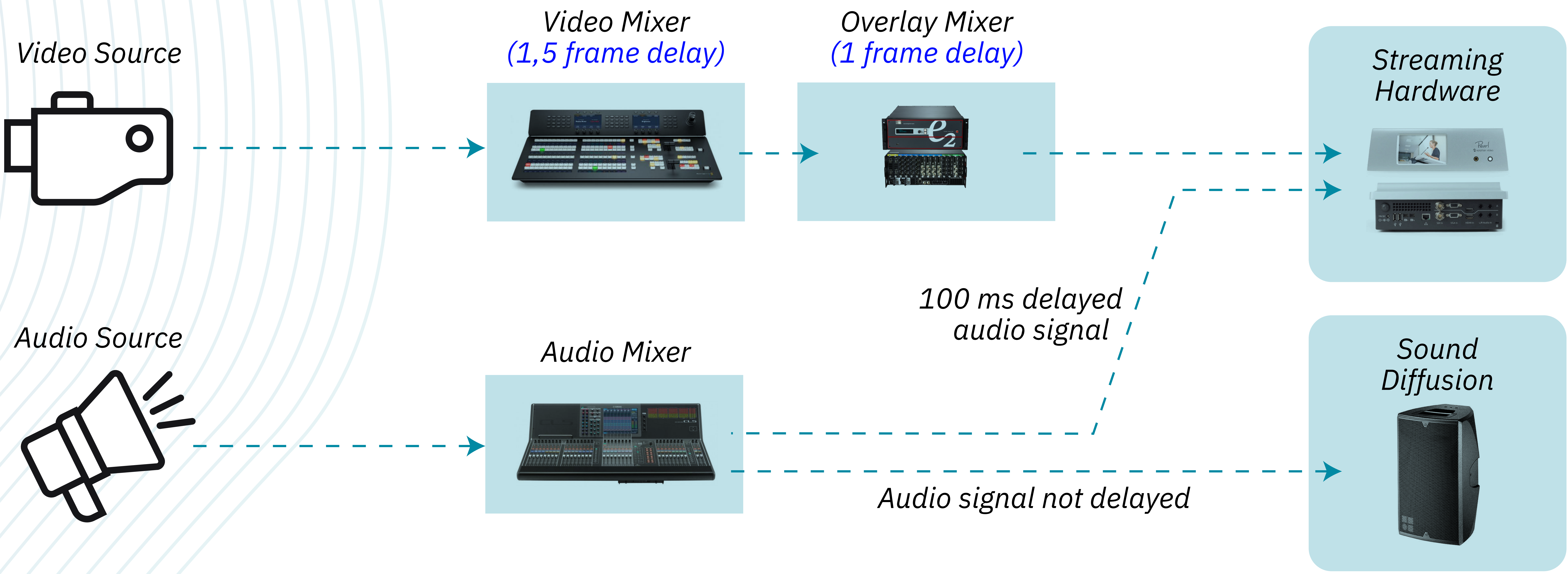
AUDIO TO VIDEO SYNCHRONISATION

Building upon the previous example where we have a video stream at 25 fps, we can determine the latency by calculating:

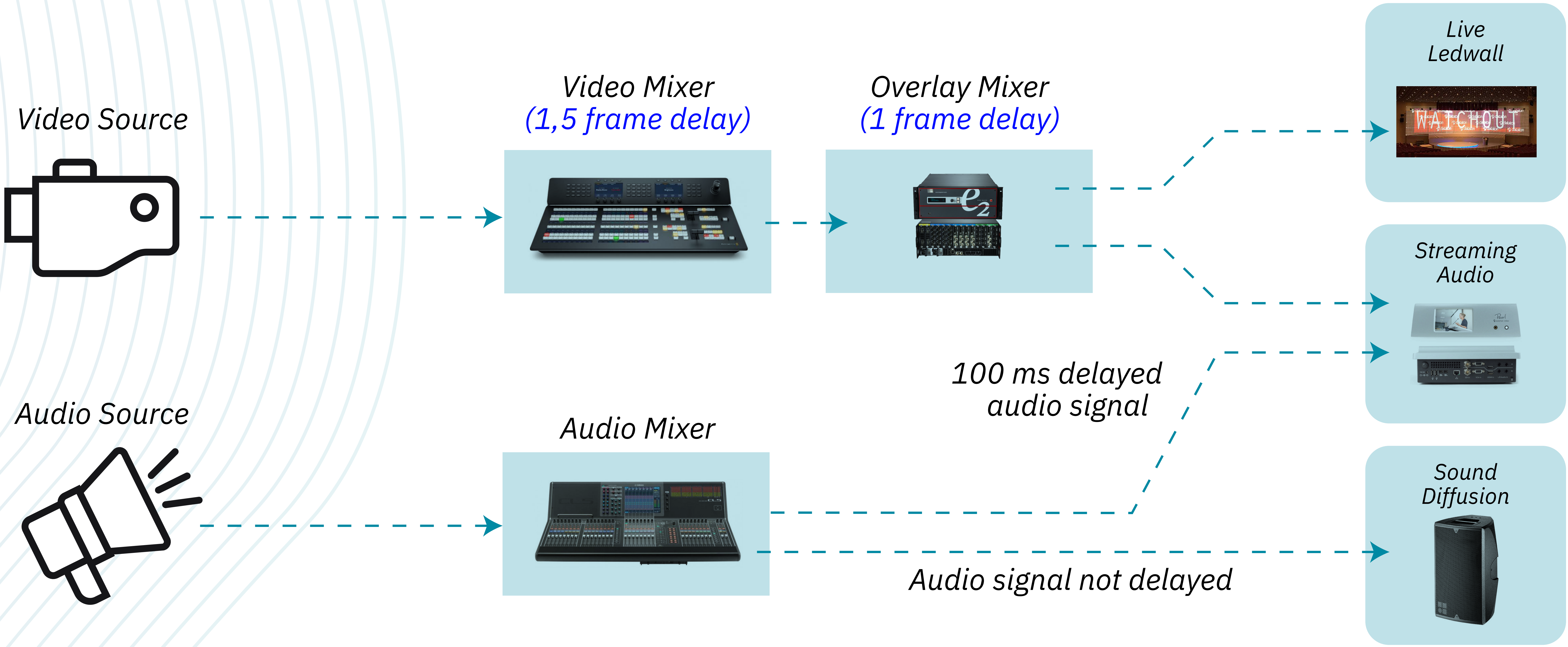
1000 ms / 25 FPS = 40 ms

However, the exact value of latency will also depend, as mentioned earlier, on the processing capabilities of the hardware used. For this reason, precise milliseconds may not be achievable, but by understanding the video hardware's frame latency, one can have a starting point, such as the 40 ms in the example.

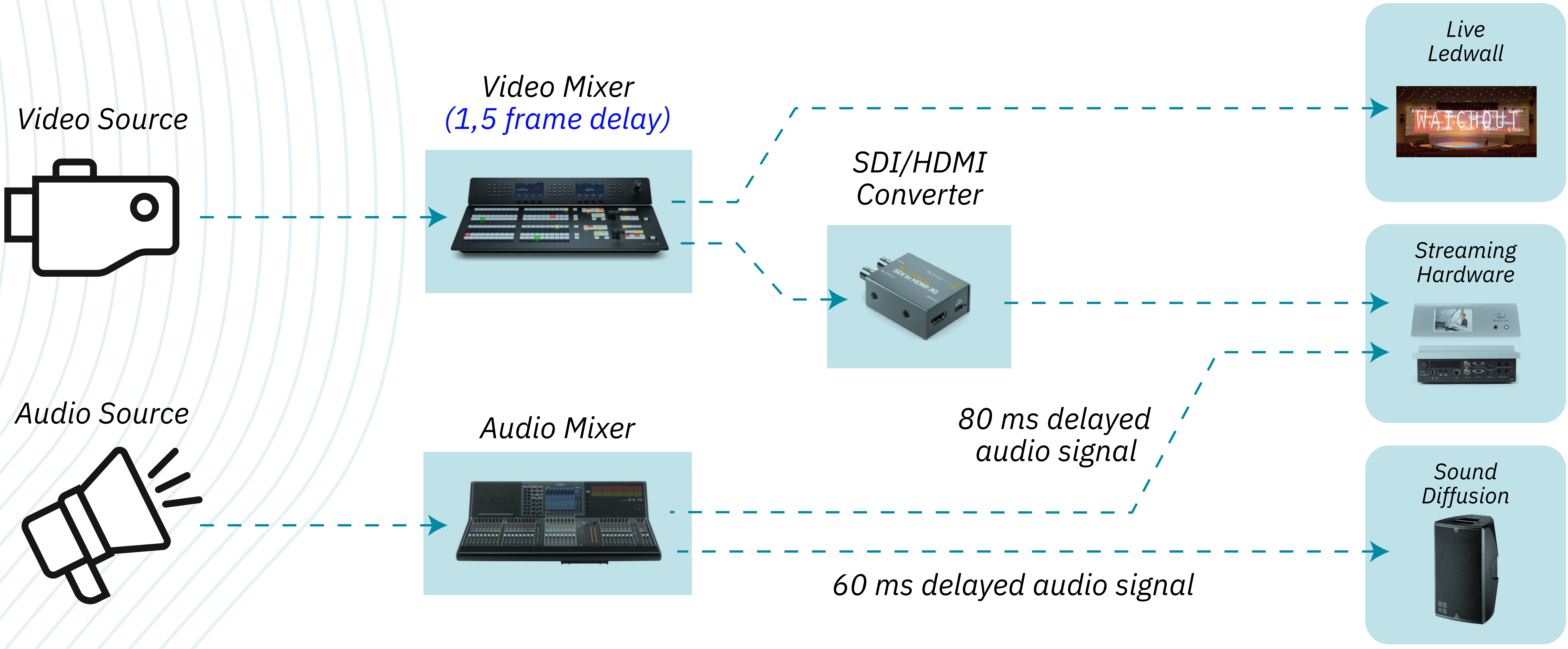
SYNCHRONISATION



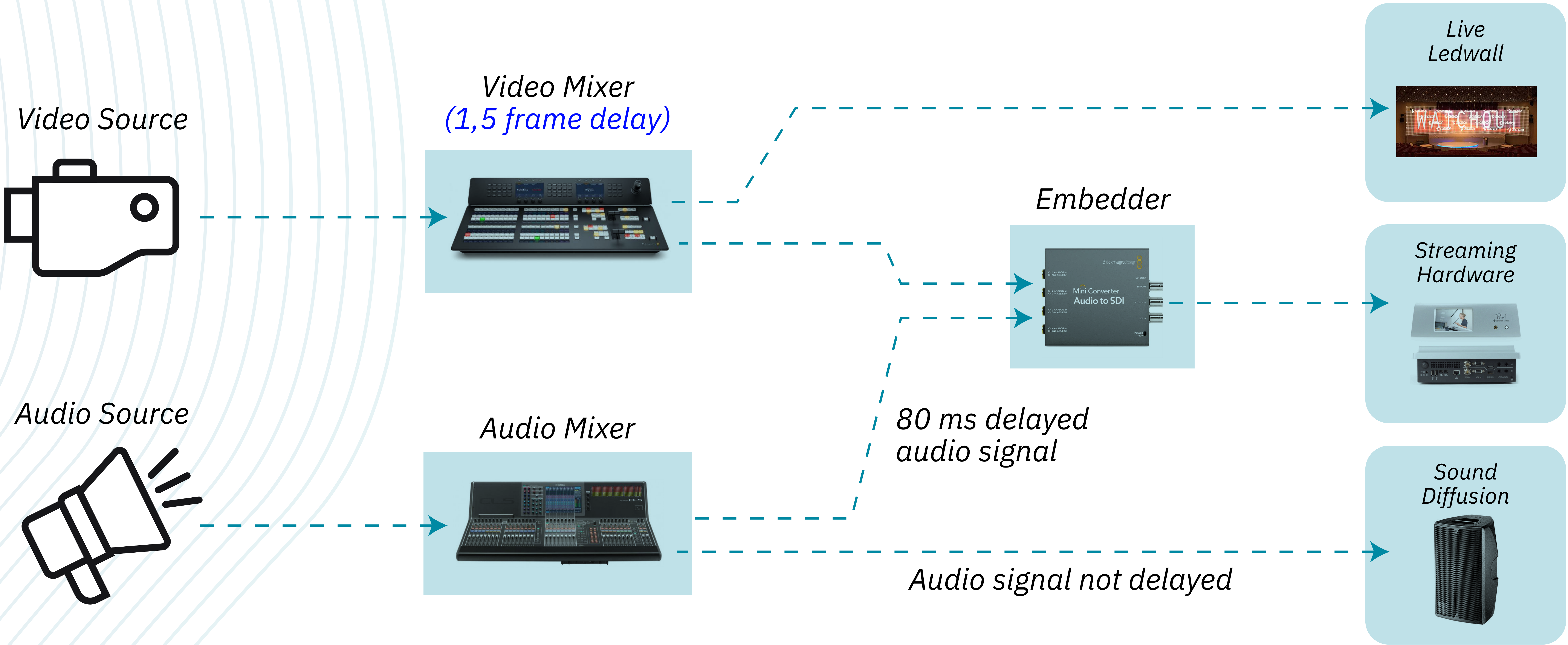
SYNCHRONISATION



SYNCHRONISATION



SYNCHRONISATION



STREAMING

The term streaming denotes the operation of sending and receiving data in a continuous way (stream) from a remote server to an end device across a network of computers.

Online streaming makes it possible for devices to play media whilst other parts of the data are being sent to the device itself.

The mechanism behind streaming involves breaking down video data into smaller chunks, called packets. These packets of data are sent to browsers, where media players then read the data as a movie.

Streaming playback begins when a device has received a sufficient number of data packets.

STANDARD WebRTC

WebRTC, Web Real-Time Communication, is the name of an open-source project which enables real-time communication for web browsers and/or mobile applications thanks to APIs (Application Programming Interfaces).
This tool provides audio/video connection on web/web-view pages, thus allowing peer-to-peer communication without a need for extra plug-ins.

Nowadays, most browsers support the WebRTC protocol.

WebRTC makes peer-to-peer communication possible. However, it does require one or more servers to function – these machines manage the various communication steps, allow clients to exchange metadata, coordinate communication between clients, and offer firewall protection.

DIFFERENCES BETWEEN STREAMING & WebRTC

GoToMeeting, Zoom, and other popular virtual meeting applications are video-conferencing platforms where (mostly) companies host webinars and online meetings. They do not rely on streaming technology – they use webRTC.

- **Streaming** – only one stream is transmitted, higher quality, transfer latency.
- **WebRTC** – multiple streams transmitted at the same time, lower quality, real-time transmission.

Streaming is employed for one-way transmission to an audience, while WebRTC is primarily used in scenarios where multiple individuals want to transmit signals, and the members of this audience may also want to transmit signals in turn.

SENDERS AND RECEIVERS

Thanks to technological advancements and telecommunications networks, live streaming is experiencing significant growth. Live streaming involves the real-time consumption of content through streaming technology.

With the exception of playback devices like audio/video streamers, today's computers, smartphones, and tablets can function as both broadcasting and receiving devices within a streaming workflow.

For instance, a live stream on platforms like YouTube, Facebook, or Twitch can be captured using a smartphone, transmitted online via the internet, and received and played back by other smartphones, computers, tablets, or streaming devices.

BUFFERING

One of the most important parts of streaming processes is buffering.

This phenomenon occurs when audio or video data is preloaded into the memory cache of a media player (buffer) during playback of online content.

Buffering allows transmission to continue even in the presence of network errors or connection issues, as the data already stashed in the buffer can still be transmitted.

However, if the internet connection is too slow, the playback device may become stuck in a buffering state – and content stored in the buffer may potentially have already been played. In such cases, it may be necessary to wait anywhere from a few seconds up to minutes until a sufficient amount of content accumulates in the buffer before playback can be resumed.

FILE RESOLUTION AND REQUIREMENTS

In streaming, the element that requires a higher data transfer rate per second is video content, whereas audio channels need fewer resources.

It is our responsibility to request enough bandwidth to be able to transmit videos at the desired resolution.

However, there may be situations where the required bandwidth is not available. In such cases, we may need to reduce the quality and, consequently, the resolution of the video content that we want to transmit. Alternatively, it may be necessary to reduce the quality of the streamed content on the receiving end in cases where the recipient's internet connection is sluggish.

FILE RESOLUTION AND REQUIREMENTS - VIDEO

SD Video Quality 720x576

For streaming, an average bandwidth of 1 Mbps to 3 Mbps is required. Image quality is significantly compromised at lower transmission speeds. At lower bandwidth values, playback interruptions can become frequent; difficulties may also arise when multiple devices are connected to the same network.

Full HD Video Quality 1920x1080

Requirements for this format range from a minimum of 4.5 Mbps to a maximum of 10 Mbps, with an average of around 6.25 Mbps.

FILE RESOLUTION AND REQUIREMENTS - VIDEO

Ultra HD or 4K 3480X2160

Ultra-High Definition (4K) requires a minimum network speed of 7 Mbps, although ultra-fast connections (25 Mbps) may sometimes be necessary. The average connection speed needed for this video resolution is equal to 19.31 Mbps.

An example is YouTube Premium, which supports 4K streaming but requires a connection of at least 20 Mbps.

FILE RESOLUTION AND REQUIREMENTS - VIDEO

In order to calculate the size of an audio file, it is necessary to consider the following parameters:

- *Sampling frequency in Hertz*
- *Bit depth*
- *Channels used*
- *Total duration of the file*

For every second of recording, we will have n samples with a certain bit depth p. The bits needed to store one second of recording will be x = np. To this value, we multiply by the number of channels used, for example, for a stereo file x2.

This way, we obtain the quantity of samples present in one second of recording.

THE STREAMING

The essential peculiarities of a means of communication for the emission of streaming content will be:

- Integrated video camera and microphone, or the possibility to connect a video camera and microphone using capture cards.
- Specialised software for the acquisition and encoding of streaming content.
- A modem with ethernet or Wi-Fi connection for the upload and transmission of content over the network.
- Another modem with ethernet or Wi-Fi connection for receiving content from the network.
- Specialised software for the decoding and playback of received content.
- A monitor and audio speaker to ensure proper playback of video content.

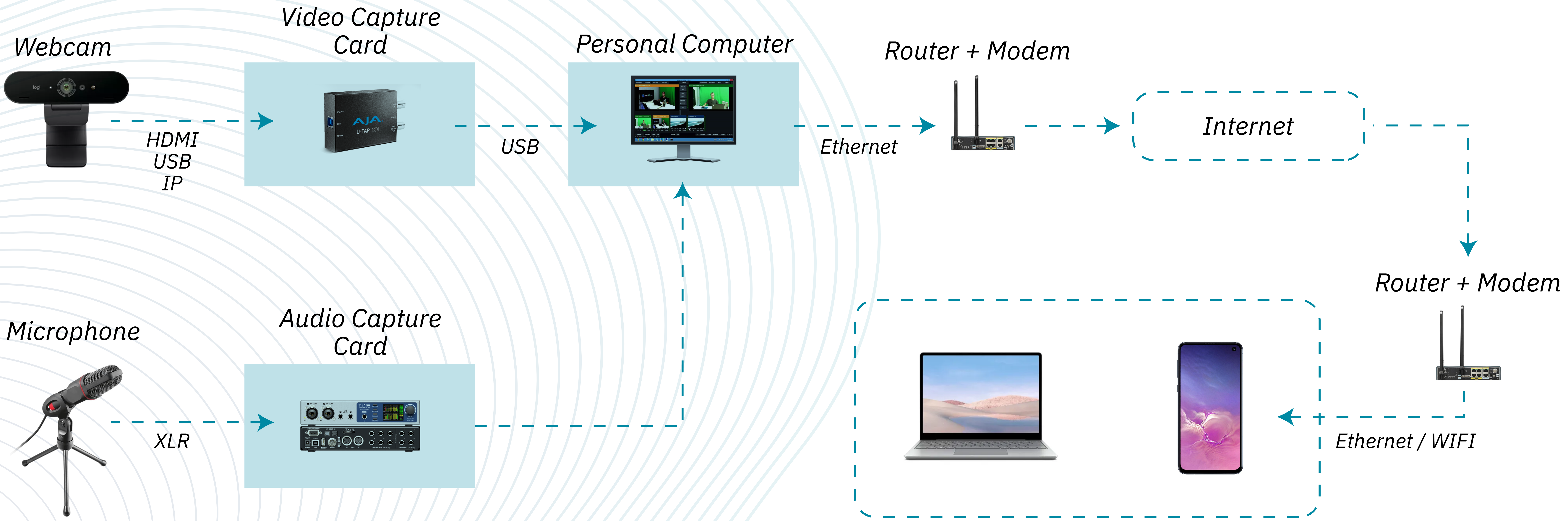
The communication devices must have enough bandwidth for a stable and high-quality transmission.

SINGLE AND DISCRETE SOURCES

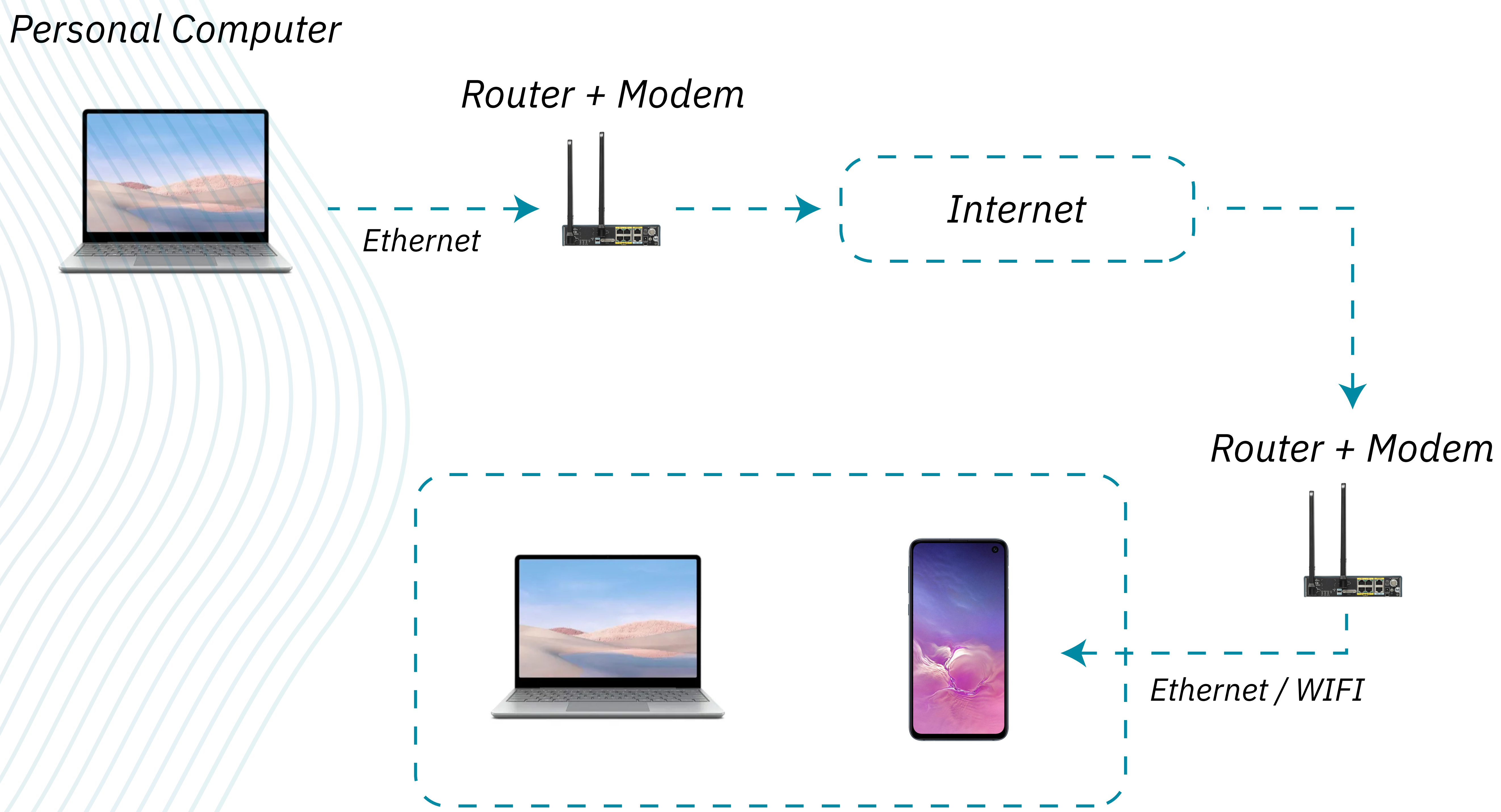
Single sources make up a category of devices designed to exclusively capture either a video or an audio signal. A webcam is a classic example of a single source.

Discrete sources, on the other hand, correspond to devices capable of simultaneously capturing both a video and an audio signal. Examples of discrete sources may include cameras with integrated microphones or computers equipped with built-in cameras and microphones.

SINGLE SOURCE STREAMING SYSTEM



DISCRETE SOURCE STREAMING SYSTEM



NOISE REMOVAL

Noise removal is the process of addressing and solving certain issues that could slow down the sending and receiving of data and thus lead to a gap in the synchronisation between audio and video channels or a less smooth playback experience.

After reaching the streaming software in the form of digital packets, correctly aligned (synchronised) audio/video files will be transmitted from the sending device to the modem for online sharing and playback.

These data packets are then processed by a receiving device, where they are temporarily stored in a buffer before being played.

A shortage of bandwidth, both during sending and receiving phases, can result in a loss of synchronisation between audio and video packets and cause the playback quality to seem choppy or jittery.

NOISE REMOVAL

For an optimal streaming experience, the use of Wi-Fi connection is not advised, as it is notoriously unreliable and particularly susceptible to unforeseeable network interruptions, interference, and sudden frequency congestions. Instead, it is recommended to use a wired connection, such as Ethernet.

It is important to ensure that all devices being utilised for the transmission and reception of data are connected via Ethernet cables conforming to and above the 5e category; categories 6 and 7 are preferred choices. These types of cables guarantee adequate data flow, e.g. 1 GB/s at 250 MHz and 10 GB/s at 600 MHz.

If wired network bandwidth is limited or unavailable, Wi-Fi solutions can be used as an emergency measure. An example of a Wi-Fi-based tool for streaming are streaming backpacks. These units, typically endowed with an Ethernet port, give users the possibility to connect to a Wi-Fi network, and also have USB ports for LTE dongles, which allow for the use of up to four sim cards.

LIVE OR ON-DEMAND STREAMING

Although the term "streaming" is often used interchangeably with "live broadcasting" for internet transmissions, it is important to clarify that a specific content delivered via streaming can actually be provided in two different modes: live or on-demand.

In the first case, the requested data is transmitted using one or more appropriate real-time compression techniques (not necessary for the transfer of pre-recorded and pre-compressed files) to lighten the load on the network and the computing units used for transmission.

The use of these specific compression techniques results in a slight latency in the transmission of the required information.

LIVE OR ON-DEMAND STREAMING

As for on-demand streaming, all audiovisual content that can be requested is already prepared for use in the form of compressed files on a server ready to fulfill requests as they come in.

Both types of streaming utilize a buffer, a small memory where data is stored before being played back.

Typical examples of on-demand streaming include Spotify and Amazon Music for music transmission, or services like Netflix, Chili, and Amazon Prime Video for the streaming of movies and TV shows.

STREAMING SOFTWARE AND HARDWARE

To conduct live streams, we can use specialised software programs that can be downloaded and installed onto a computer or we can instead use hardware designed especially for this purpose.

The two most widely used programs for audio and video signal acquisition, mixing, and broadcasting are OBS and vMix. These applications transform video signals into a video stream that can be sent to a media server.

If no computer is available or if installing software is not a viable option, it is possible to use hardware devices to obtain a video stream.
Such devices are often very specifically designed for streaming purposes; they are endowed with video camera input, a monitor output port, and an ethernet cable connection for streaming transmission

GIANLUCA PATRITO
Streaming technician