



HAL
open science

A suitable audio framework design for videoconferencing and online teaching

Roland Cahen

► **To cite this version:**

Roland Cahen. A suitable audio framework design for videoconferencing and online teaching. 2020.
hal-02542633

HAL Id: hal-02542633

<https://hal.science/hal-02542633>

Preprint submitted on 14 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A suitable audio framework design for videoconferencing and online teaching

Roland Cahen
Artistic professor and researcher
Centre de Recherche en Design
ENSCI les Ateliers – ENS Paris – Saclay
Roland.cahen@ensci.com

At the moment most software tools for videoconferencing used by teachers for their online courses are prone to having a problematic audio framework. It is easy to speak, listen and interact with the videoconference tools available. It is also possible to share visuals, a screen or an application window. But it becomes much more difficult when you want to share a sound, the sound coming from a screen, the sound of an application or even a simple video. This is especially the case when the sound client is the same as the standard OS audio client. To share sound as well as images, you have to create a virtual output bus for your application which can be received by the videoconferencing software. Then you have to switch from your microphone input to the application bus, backwards and forward, each time you either present or speak, but you can't do both at the same time. Of course, using multiple audio busses inside the computer is possible for skilful teachers with a background in audio engineering, but difficult for most other teachers and students.

Therefore, I firmly believe the videoconferencing audio architecture should be improved to make online teaching easier. This short paper tries to define sensible expectations and a possible framework design, without thoroughly going into the domain/science of audio engineering.

The audio architecture of computers is designed in such a way that, audio interconnections between applications are generally hidden to users who can only select one input and one output. However, in order to share sound and hear the videoconference at the same time, you need to split at least two inputs and access one or two separate outputs inside the videoconference software. More importantly, you need two busses for your computer outputs:

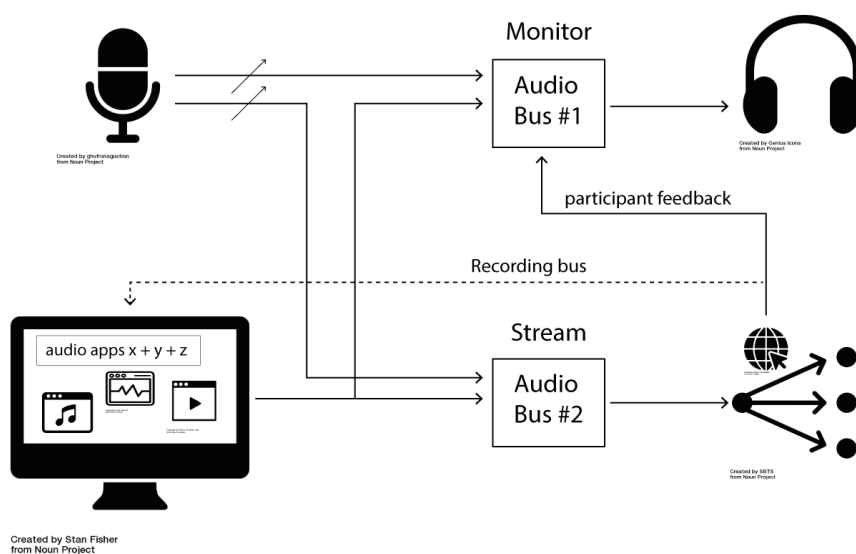
- 1) A monitoring bus to listen to your distant conversation
- 2) A stream bus to send and receive presentation sounds

And one or two for the computer inputs:

- 1) A monitoring bus for your microphone
- 2) An optional bus for recording an audio stream coming from the internet

The following illustration presents a suitable audio framework design.

Proper audio framework design for videoconferencing



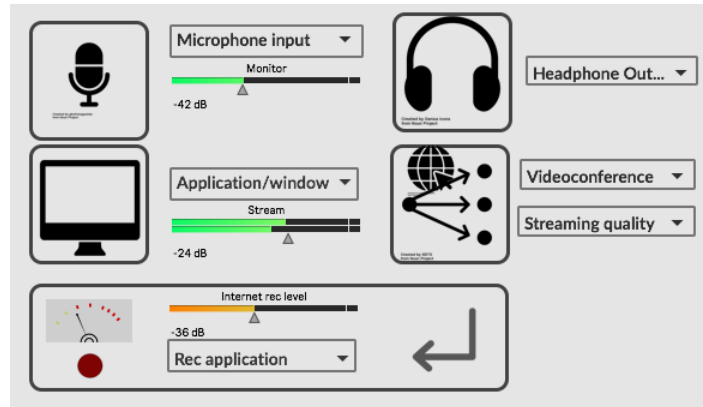
N.B. The same software unit can be used by the teacher and the student

roland.cahen@ensci.com (2020)

Generally, the microphone bus is mono and the stream bus stereo, but they can both be stereo.

The same tools can be used by teachers and students to easily exchange audio both ways. This framework could be implemented either directly into the videoconference software or in third party applications. In order to be efficient, it should be simple and adaptable to the performances of various machines, internet debit and hardware equipment.

A simple interface like the following could give access to most of the features required.



In addition, some latency issues will need to be solved. When you listen to your own voice feedback coming from your microphone, it must not be delayed. The less delay there is, the better it is. A delay of 30ms. is acceptable but not more. There are various ways to solve this issue: either by using open earphones instead of closed headphones and muting the feedback, or by plugging the microphone input signal into the headphone output as soon as possible in the audio chain. Another solution would be to reduce the input / signal buffer sizes and give priority to audio on other computer tasks.

When you teach the use of a video or a sound software for example, you also want your presentation to stream without too much latency and with a good synchronisation between audio and visual actions. A delay of 1 sec. between emission and reception and an audio/visual desynchronization of 50 ms. are acceptable.

I believe the modification of the audio architecture and the setting up of a flexible audio interface such as shown in this introduction could thoroughly improve teaching virtually and using videoconference tools. I would very much like to explore solutions to implement this design, because I think videoconferencing will become an essential tool within education in the future.

Roland Cahen 13/04/2020

C
E N T R E
DE RECHERCHE EN
D E S I G N
ENSCI - LES ATELIERS
ÉCOLE NORMALE
S U P É R I E U R E
P A R I S
S A C L A
Y