

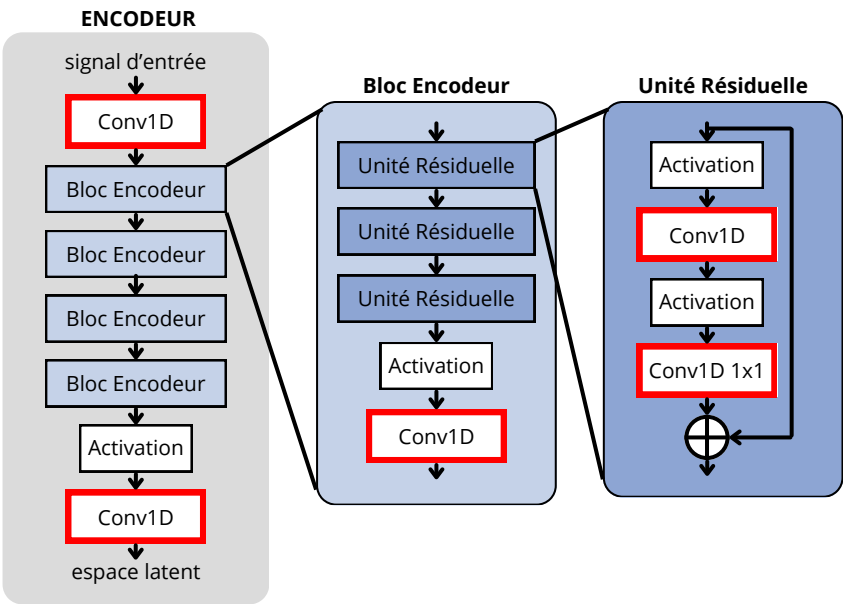
COMPRESSION AUDIO PAR DEEP LEARNING

Impacts des convolutions causales

En téléphonie mobile, la compression audio est réalisée par un codec. A l'émission, le codec permet de réduire la taille de l'information utile qui sera transmise dans le réseau. A la réception, le codec réalise le travail inverse. L'intelligence artificielle commence à remplacer les codecs classiques par des codecs audio neuronaux. Bien que leur usage soit encore limité, Google utilise Lyra V2 [1] pour Google Duo, et Microsoft a intégré un codec neuronal appelé Satin [2] dans Teams. Le codec DAC [3] (Descript Audio Codec) est à ce jour le codec le plus performant de l'état de l'art. Toutefois, DAC n'est pas adapté aux cas d'utilisations en temps réels comme la téléphonie.

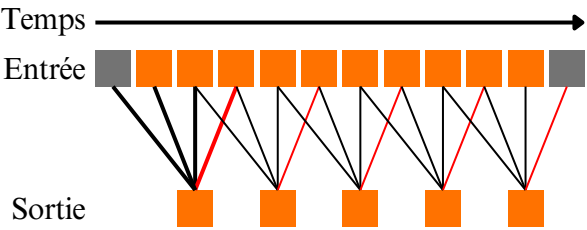
Couches du réseau encodeur de DAC

L'encodeur de DAC utilise une succession de couches pour faire passer le signal d'entrée dans un espace latent de dimension plus faible. Dans DAC, l'utilisation des convolutions **non-causales** rend impossible son utilisation en temps réel.

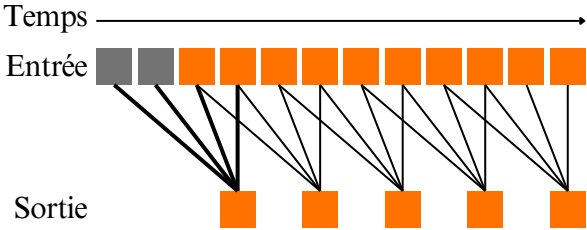


Les convolutions dans un réseau de neurones peuvent permettre de réduire la taille de l'information utile. Il existe des convolutions **causales** et **non-causales**. D'un point de vue temporel, la convolution non-causale utilise des échantillons "futurs" pour calculer l'élément à un instant t.

Convolution Non-Causale



Convolution Causale

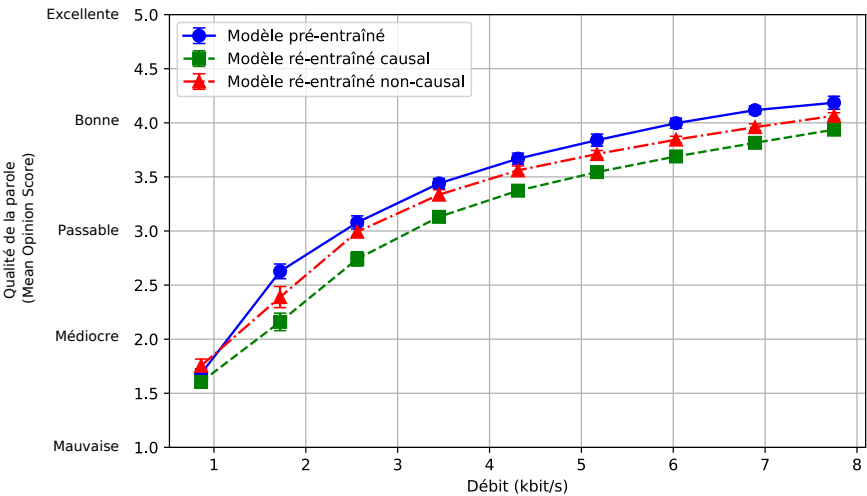


Modification de Descript Audio Codec

Nous avons mis en place une version causale de Descript Audio Codec. Dans cette version, la causalité nous permet de supprimer le retard induit par l'attente d'échantillons futurs (ce retard rend DAC inadapté au cas d'utilisation en temps réel). Toutefois, la suppression du contexte "futur" réduit les performance à un débit donné.

Analyse de la qualité via notation objective (PESQ↑)

Pour Descript Audio Codec, appliquer des convolutions causales dans le réseau de neurones réduit la qualité des signaux audio décodés.



Références :
[1] Neil Zeghidour et al. *SoundStream: An End-to-End Neural Audio Codec*. 2021.
[2] Jigar Dani et al. *Satin: Microsoft's latest AI-powered audio codec for real-time communications*. 2022.
[3] Rithesh Kumar et al. *High-Fidelity Audio Compression with Improved RVQGAN*. 2023.