

**VOXYGEN | SERVER**  
**Fiche technique**  
*Version : 9.4*

**Référence** : VOX351\_SERVER\_Datasheet\_9.4\_1.1\_FR

**Version** : 1.1

**Statut** : Release

**Date** : 04/06/2024

**Diffusion** : Restricted



## Sommaire

<b>1</b>	<b>PRÉSENTATION</b>	<b>3</b>
<b>2</b>	<b>CONFIGURATION REQUISE</b>	<b>4</b>
2.1	Système d'exploitation requis	4
2.2	Exigences matérielles	4
<b>3</b>	<b>ENTRÉES / SORTIES</b>	<b>4</b>
3.1	Entrées	4
3.2	Sortie	4
<b>4</b>	<b>PROTECTION CONTRE LA COPIE</b>	<b>5</b>
<b>5</b>	<b>LIMITES</b>	<b>5</b>
<b>6</b>	<b>PERFORMANCE</b>	<b>5</b>
<b>7</b>	<b>AUTRES DOCUMENTATIONS</b>	<b>5</b>



# 1 Présentation

Voxygen SERVER est une solution de synthèse vocale à déployer dans votre infrastructure.

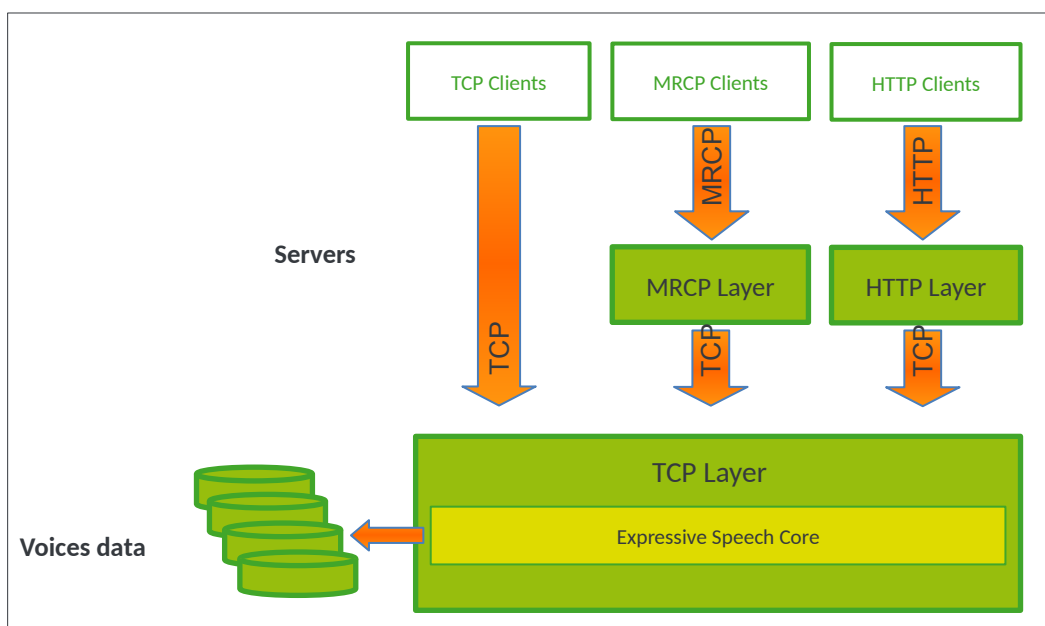
Voxygen SERVER intègre le moteur de synthèse vocale Voxygen CORE. Celui-ci permet la création en temps réel de plusieurs flux vocaux simultanés. Ces flux vocaux appelés «ports TTS» offrent des temps de réponse très rapides et maîtrisés.

Voxygen SERVER est multilingue, rapide, robuste et évolutif. Il permet de traiter des dizaines à des milliers d'interactions vocales simultanées. Il est facile à intégrer via des interfaces standardisées telles que MRCP et HTTP. Il respecte les normes W3C : SSML et PLS.

L'application cliente envoie directement le texte au Voxygen SERVER. Des paramètres optionnels permettent par exemple, de choisir une voix spécifique pour lire le texte, ou de choisir le format du flux audio. Le flux audio est généré progressivement par Voxygen SERVER au fur et à mesure du traitement du texte d'entrée.

Trois interfaces sont disponibles pour interagir avec le Voxygen SERVER. Ces interfaces sont appelées "Layers" :

- **HTTP Layer** est un serveur HTTP léger qui offre un service de synthèse simple via une API "REST-like". L'appel du service se fait avec une simple requête HTTP GET ou POST avec un ensemble de paramètres requis. Le serveur n'établit pas de sessions avec les clients. Le service est conçu pour fonctionner sur un réseau privé, aucune authentification n'est donc fournie.
- **MRCP Layer** offre une interface standardisée basée sur les protocoles W3C MRCP v1 et v2.
- **TCP Layer** est un serveur léger basé sur des sockets TCP. Le protocole propriétaire est simple et facile à mettre en œuvre. Des exemples d'implémentations clientes sont fournis en C, java et Python.





## 2 Configuration requise

### 2.1 Système d'exploitation requis

- Linux RHEL 8, RHEL 9 (ou équivalent version Rocky)
- Linux DEBIAN 11, 12
- Linux UBUNTU 22.04, 24.04
- Windows Server 2016, 2019, 2022

### 2.2 Exigences matérielles

- **Processeur :**
  - architecture x86\_64
  - les flags AVX2 et FMA doivent être disponibles sur les CPU.
  - chaque cœur-CPU physique doit disposer d'au moins 2 SMT threads. Une paire de SMT threads est nécessaire pour qu'un flux d'audio soit rendu en temps réel.
- **Espace disque :**
  - 80 MB pour les exécutables Voxygen SERVER
  - 100 à 1700 MB par voix
  - 50 MB pour la gestion des logs
  - X MB pour les autres logs (dépend de la configuration réalisée)
- **Mémoire RAM :**

Le besoin en mémoire RAM est la somme de deux valeurs :

  - 1/ Mémoire RAM pour les ports TTS : 30 Mo de RAM par port TTS configuré.
  - 2/ Mémoire RAM pour la mise en œuvre des voix : au moins la taille de la plus grosse voix installée. Pour des performances optimales, la taille de la mémoire pour les voix doit être suffisante pour charger toutes les voix utilisées en mémoire.

Exemple: pour fonctionner avec des performances optimales, un serveur TTS avec 200 ports TTS et 2 voix a besoin de 8 Go de RAM.

## 3 Entrées / Sorties

### 3.1 Entrées

Voxygen SERVER est conforme au langage standard de balisage de synthèse vocale du W3C (SSML versions 1.0 et 1.1). Le texte à vocaliser peut être un texte brut encodé en UTF-8 ou un document SSML. Voxygen SERVER est également conforme à la spécification du W3C pour les lexiques de prononciation (PLS version 1.0).



## 3.2 Sortie

La sortie du signal vocal peut être échantillonnée à n'importe quelle fréquence entre 6 kHz et 48 kHz. Les échantillons sont numérisés sur 16 bits et peuvent être compressés en G.711 (8 bits, loi A ou loi mu). Le signal vocal peut être généré au format brut sans en-tête ou au format WAV. La compression MP3 et Ogg / Vorbis est également possible, à condition qu'un logiciel tiers adéquat (bibliothèque Lame pour MP3 et Xiph pour Ogg / Vorbis) soit disponible et installé sur la machine hôte (non disponible pour MRCP Layer). Le signal vocal peut être mélangé avec des fichiers audio externes. Le signal vocal est généré à la volée, au fur et à mesure du traitement du texte source.

## 4 Protection contre la copie

Voxygen SERVER est protégé contre la copie. Son utilisation sur un ordinateur nécessite la présence d'un fichier d'activation à installer à côté du logiciel. Un fichier d'activation lie une version donnée du Voxygen SERVER à sa machine hôte.

## 5 Limites

Une instance du Layer TCP ne peut pas traiter plus de 200 ports TTS. Cependant il est possible de déployer plusieurs instances sur un ordinateur.

Le format de sortie du Layer MRCP est limité aux valeurs suivantes:

- G.711 compressé (A-law ou mu-law, 8 bit) à 8kHz.
- 16 bits PCM à 8 ou 16 KHz.

## 6 Performance

Sur un processeur monocœur i5 x64 3GHz, Voxygen SERVER est capable de gérer jusqu'à 80 ports TTS simultanément en temps réel. Les performances peuvent être moindres lorsqu'un signal vocal avec en-tête est requis (wav, mp3 ou ogg) ou lorsque le signal vocal est compressé.

## 7 Autres documentations

*VOX171\_SERVER\_presentation*  
*VOX62\_syntserver\_reference\_manual*  
*VOX342\_syntHTTPserver\_reference\_manual*  
*VOX122\_MRCP\_reference\_manual*  
*VOX300\_UNIMRCP\_reference\_manual*  
*VOX170\_exceptions\_lexicon\_manual*  
*VOX341\_Text\_normaliser\_manual*  
*VOX31\_SSML\_reference\_manual*  
*VOX32\_Baratino\_tags\_reference\_manual*

