

PROJET STATION TV



CAHIER DES CHARGES

SUIVI DES MODIFICATIONS DU DOCUMENT			
Version	Date	Validation	Commentaire
1.0	08/10/2025	Frédéric Chauvin	Rédaction initiale

REDACTEUR : Dorian BRISSON	VALIDATEUR : Frédéric CHAUVIN
CLIENT : Polytech Tours	ENCADRANT : Mathieu DELALANDRE

OBJET DU DOCUMENT

Le présent document a pour objectif de définir le cahier des charges fonctionnel et technique du projet « Transcription audio haute performance sur Station TV », mené dans le cadre du module Projet de Recherche et d’Innovation (PRI) de 5^e année à Polytech Tours (filière ISIE). Ce projet s’inscrit dans la continuité du programme de recherche Station TV, initié au LIFAT (Laboratoire d’Informatique Fondamentale et Appliquée de Tours). Il vise à optimiser la chaîne de transcription automatique audio vers texte (Speech-To-Text, STT) basée sur l’outil Whisper (OpenAI) afin de permettre une transcription rapide et scalable des flux télévisuels issus de la TNT.

Le projet comprend une phase de mise à niveau matérielle (installation de 256 Go de mémoire ECC sur la station Dell 5820), une phase de benchmark et d’optimisation logicielle, ainsi qu’une phase d’industrialisation du pipeline dans l’infrastructure Station TV.

FONCTIONNEMENT DÉSIRÉ

Le système développé doit permettre la transcription automatique de flux audio complexes issus des chaînes de télévision capturée par la Station TV.

Le fonctionnement attendu suit les étapes suivantes :

1. Initialisation du système :
 - Lancement des scripts de traitement et chargement des modèles Whisper.
 - Vérification automatique de la configuration matérielle (RAM, CPU, stockage).
2. Préparation des données audio :
 - Lecture des fichiers audio capturés (format MP3/256 kbps).
 - Découpage automatique en lots (batches) selon un protocole temporel défini.
3. Transcription :
 - Sélection du modèle Whisper (small, medium, large).
 - Exécution parallèle multi-process (jusqu'à 36 threads sur CPU Xeon W-2295).
 - Production des transcriptions textuelles au format .txt ou .csv.
4. Évaluation et métriques QoS :
 - Mesure du temps de traitement, du débit (throughput), de l'utilisation CPU/RAM.
 - Comparaison de la qualité linguistique (taux d'erreur WER, stabilité des segments).
5. Analyse et visualisation :
 - Génération automatique de rapports et graphiques de performance.
 - Export des résultats et intégration dans le pipeline Station TV.
6. Maintenance et évolutivité :
 - Documentation complète (installation, scripts, résultats).

L'EXISTANT

Infrastructure matérielle :

- Machine Dell Precision 5820 – Xeon W-2295 (18 cœurs / 36 threads).
- 64 Go DDR4 ECC installés avant mise à niveau.
- Station complémentaire Dell PowerEdge T640 (80 threads).
- Stockage total : 38 To (RAID interne) + 190 To externe.

Infrastructure logicielle :

- OS : Windows Server 2022 / Ubuntu 22.04.
- Librairies : Whisper (OpenAI), Torch, FFmpeg, NumPy, Pandas.
- Scripts existants : BasicTestWhisper.py, RunBatchWhisper.py, ComputeQoS.py.

Bases de données audio :

- Hello World : 588 heures audio issues de 7 chaînes TNT.
- Bases synthétiques de test : baseMod, baseBatch, baseMtr.

Constat :

- Saturation mémoire sur les modèles medium et large.
- Trashing CPU lors d'exécutions parallèles intensives.
- Absence de métriques QoS centralisées.
- RAM insuffisante pour les transcriptions longues à haute qualité.

CAHIER DES CHARGES

Objectifs principaux

- Permettre l'installation et la validation de 256 Go de RAM ECC sur la Dell 5820.
- Garantir la compatibilité BIOS, la stabilité et la détection quad-channel.
- Réaliser des tests de performance Whisper sur les modèles *small*, *medium* et *large*.
- Mettre en place un système d'évaluation QoS automatisé.
- Industrialiser le pipeline de transcription audio dans l'écosystème Station TV.

Fonctionnel

Matériel :

- Installation supervisée par la DSI Polytech.
- Validation complète post-installation (stress test RAM/CPU).

Logiciel :

- Script principal capable de transcrire des fichiers de plusieurs heures.
- Gestion des fichiers batch/run.
- Sauvegarde automatique des résultats et logs.

Performance :

- Objectif : traitement de 588h d'audio en < 12h (x50 temps réel).
- Débit minimal requis : 5× pour *small*, 1× pour *medium*, 0.5× pour *large*.
- Mémoire utilisée : ≤ 240 Go (sans swapping).

Qualité de transcription :

- Word Error Rate (WER) ≤ 10 % sur segments clairs.
- Absence de coupures ou erreurs de segmentation.