
Transcription audio haute performance sur Station TV

PRI 5A - Formation ISIE
Polytech Tours - 2025-2026

Chef de projet : Dorian BRISSON
Client : M. DELALANDRE

Sommaire

- Contexte et existant
- Problématique et objectifs
- Gestion de projet
- Choix techniques
- Conception
- Protocole expérimental
- Résultats
- Conclusion et perspectives



Contexte et existant

Contexte : Station TV

- Projet Station TV (LIFAT)
- Flux TNT
- Plateforme de calcul parallèle
- Services : analytics TV,
fact-checking, détection
publicité..



- Transcription
audio



Matériel

- Dell Precision 5820
- CPU : Intel Xeon W-2295 (18 cœurs physiques, 36 threads)
- RAM : 64 Go
- OS : Windows



Problématique et objectifs

Problématique

- Audio TNT complexe
- Volumes importants
- Qualité vs performance
- Stabilité long run

Existant : travaux précédents

- Transcriptions possible
- Étude STT existante (speech to text)
- Choix de Whisper
- Script multi-processus

Limites de l'existant et constats

- Metrics fait main
- Peu d'industrialisation
- Rechargement fréquent du modèle
- Limitation mémoire identifiée

Objectifs du projet

- Augmenter la RAM de 64Go à 256Go
- Mise en place de metrics
- Tests sur les hauts modèles medium/large
- Pipeline run/batch



Gestion de projet

Organisation

Rôle	Nom	Organisation
Chef de projet	Dorian BRISSON	Polytech Tours
Client / Encadrant	Mathieu DELALANDRE	Polytech Tours

Méthodologie

- Méthodologie agile
- Travail par sprints





Choix techniques

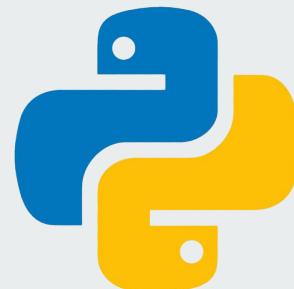
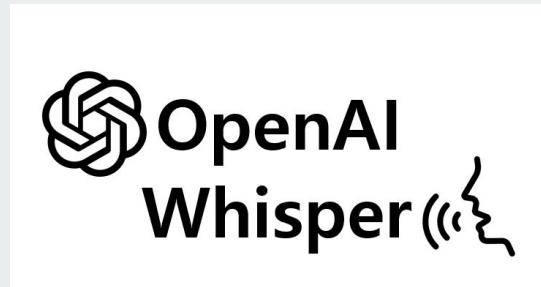
Outils de transcriptions

Outil / Variante	WER (FR) approximatif	Vitesse CPU	RAM nécessaire typique (quantité)
Whisper (full)	~8-15 %	Moyen	~2-5 GB+
Vosk (modèles FR)	~10-30 %	Très rapide	50-500 MB modèles légers
wav2vec2-fr (base)	~12-20 %	Moyen	~1-2 GB
Kaldi avec modèle FR	~15 %+	Très rapide	~100-600 MB (dépend modèle)
Coqui STT / DeepSpeech (FR)	~15 %+	Moyen à rapide	~500 MB-1 GB

Source : northflank.com

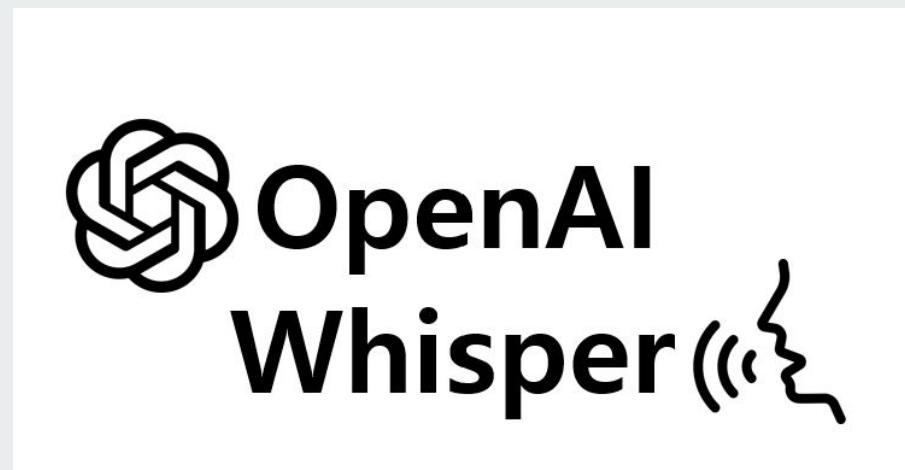
Technologie retenu

- Whisper
- Python



Whisper : principe général

- Modèle STT
- Architecture Transformer
- Multilingue
- Robuste au bruit



Whisper : modèles et compromis

- tiny, base, small, medium et large
- Qualité croissante
- Coût RAM croissant

Défis système liés à Whisper

- Forte consommation RAM
- Temps de calcul
- Parallélisation nécessaire



Conception

RAM

TICKET n°100769 : Installation RAM haute performance - station TV ACTIONS ▾

Statut :	Archivé
Catégorie :	DSI - Acquérir du matériel informatique
Propriétaire :	Mathieu Delalandre (delalandre)
Gestionnaire :	Sebastien Beaufils (sbeaufils)

[Historique](#) [Propriétés](#) [Fichiers attachés](#) [Invités](#)

13/10/25 16:01 : le ticket a expiré.
 29/09/25 15:13 : Sébastien Beaufils (sbeaufils) a clos le ticket.
 RAM installée, test -> OK
 29/09/25 15:13 : Sébastien Beaufils (sbeaufils) a modifié le temps passé.
 24/09/25 17:32 : Dorian Brisson (21801799) a ajouté un commentaire.
 Je serai présent. Je vais prévenir mon enseignant par précaution.
 Bien cordialement,
 24/09/25 17:29 : Mathieu Delalandre (delalandre) a ajouté un commentaire.
 Oui c'est ok, merci Dorian de vous rendre disponible et de prévenir au besoin votre enseignant de 14h (si retard en cours).
 Bien Cordialement,
 24/09/25 09:53 : Sébastien Beaufils (sbeaufils) a ajouté un commentaire.
 Lundi 13h30 ?
 24/09/25 09:36 : Mathieu Delalandre (delalandre) a ajouté un commentaire.
 Plus lundi, de préférence début d'après-midi ou à partir de 9h30 le matin.
 Devons nous prévoir de déplacer la machine, ou on pourra faire cela directement sur la station.
 Je reprécise qui y a deux jeux de barrette, soit 4x 64 Go.
 Merci
 24/09/25 09:13 : Sébastien Beaufils (sbeaufils) a pris le ticket en charge (précédemment libre).
 Vendredi ou lundi prochain ?
 23/09/25 15:30 : Mathieu Delalandre (delalandre) a ajouté Dorian Brisson (21801799) à la liste des invités.
 23/09/25 15:28 : Mathieu Delalandre (delalandre) a ajouté Sébastien Beaufils (sbeaufils) à la liste des invités.
 23/09/25 15:28 : Mathieu Delalandre (delalandre) a créé le ticket.

Panier – 1 Produit(s)

Vendu(s) par : Conrad Electronic

Kingston KTD-PE432/64G Module mémoire pour PC DDR4 64 GB 1 x 64 GB ECC 3200 MHz DIMM 288 broches CL22 KTD-PE432/64G

Livraison : 04.07.2025 - 08.07.2025

Code produit : 2823381

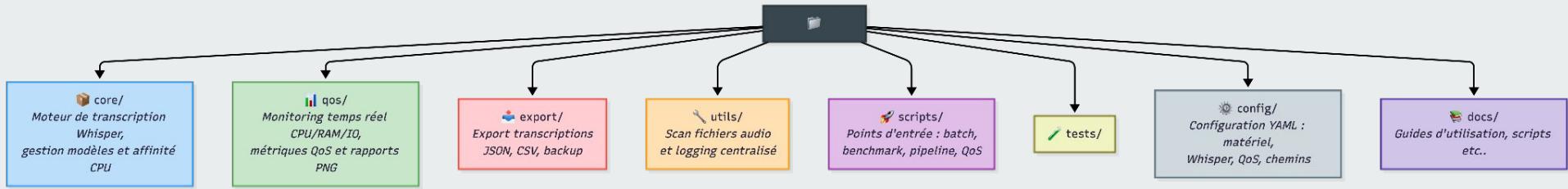
– pièces 2 +

652,50 €
Avec 10% de remise client
Prix unitaire 326,25 €

Prix HT	652,50 €
Frais d'expédition	8,99 €
TVA	132,30 €
Prix total TVA incluse.	793,79 €

Architecture globale du système

- Approche modulaire
- Chaque module est indépendant
- Maintenabilité du code



Fichier de configuration

Hardware — Matériel cible

- Nombre de coeurs physiques
- Nombre de threads logiques
- Nombre maximal de processus Whisper simultanés
- Quantité totale de mémoire vive
- Seuil maximal d'utilisation de la RAM

Moteur de transcription

- Modèle Whisper utilisé
- Langue de transcription
- Type de périphérique d'exécution (CPU / GPU)
- Affinité CPU par processus
- Formats de sortie générés (TXT, SRT)
- Horodatage au niveau des mots

Fichier de configuration

QoS — Supervision & monitoring

- Intervalle de mesure
- Mesures d'utilisation CPU
- Mesures d'utilisation mémoire
- Mesures d'entrées / sorties disque
- Seuil d'alerte CPU (avertissement)
- Seuil d'alerte CPU (critique)
- Seuil d'alerte RAM (avertissement)
- Seuil d'alerte RAM (critique)
- Suivi de la consommation énergétique
- Estimation du coût énergétique

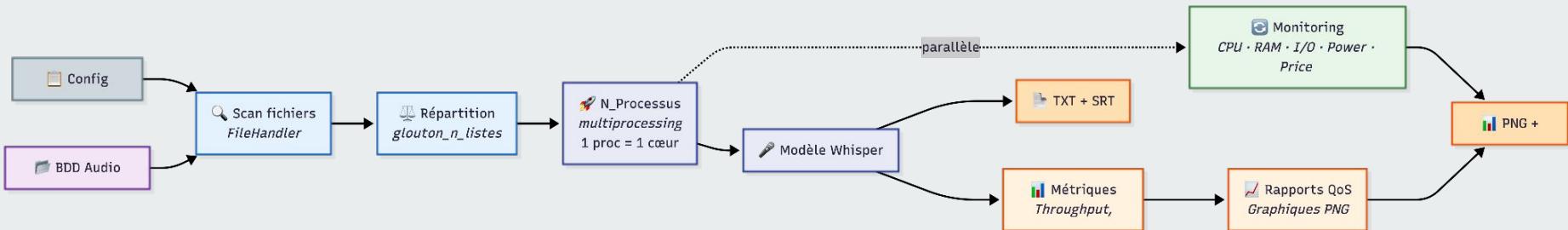
Export — Sortie des résultats

- Encodage des fichiers de sortie
- Format des métadonnées
- Politique de sauvegarde automatique
- Fréquence des sauvegardes
- Convention de nommage / préfixe d'indexation

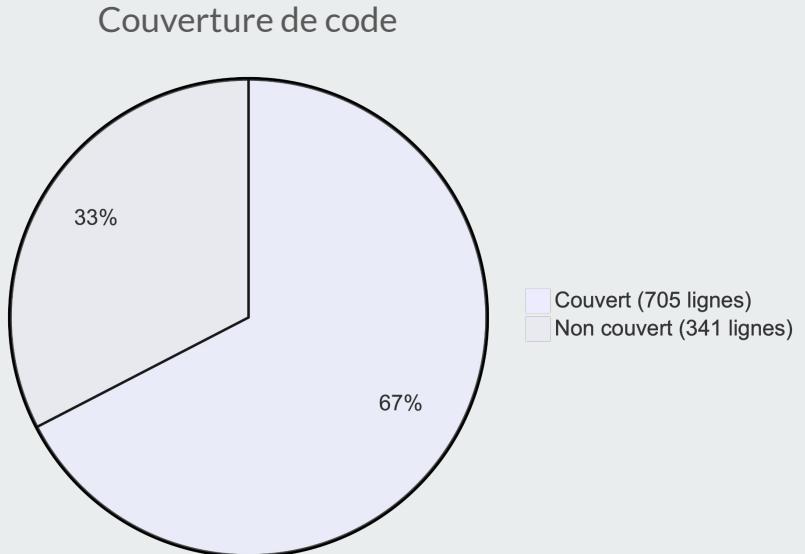
Paths — Chemins d'accès

- Répertoire d'entrée des fichiers audio
- Répertoire de sortie des transcriptions
- Répertoire des rapports
- Répertoire des journaux (logs)
- Répertoire des métriques et trackers
- Répertoire des sauvegardes

Pipeline



Tests unitaires



Metrics

- Taux d'utilisation CPU en %
- Charge mémoire en GB
- Consommation en kWh
- Prix
- Troughput
- Temps moyen pour chaque processus
- I/O
- WER

Apports techniques majeurs

- Meilleures performances
- Plus de metrics et intégré
- Pipeline industrialisé :
 - Fichier de configuration
 - Lancement avec un .bat



Protocole expérimental

Bases de test utilisées

- BaseMod 5 fichiers : 0, 240, 480, 720, 960, 1200 en secondes
- BaseBatch : 32 dossiers qui contiennent jusqu'à 24 fichiers d'1h d'audio.
- Base HelloWorld (cible)

Benchmarks

- Benchmark 10×5×modèles
- tiny → large
- Mesures throughput

	tiny	base	small	medium	large
0	0	0	0	0	
240	55,35	97,04	270,48	778,76	
480	97,55	176,23	545,26	1435,22	
720	183,9	317,53	752,3	2181,12	
960	264,14	378,22	1311,25	2861,25	
1200	301,51	478,76	1420,1	3630,68	
Th	3,980	2,506	0,845	0,331	
1/Th	0,251	0,399	1,183	3,026	



40 à 50% de gain

	tiny	base	small	medium	large
0	0	0	0	0	0
240	39,66	65,22	176,39	526,06	1136,98
480	76,69	155,73	405,5	1038,24	1822,16
720	127,74	198,57	626,42	1503,48	2694,27
960	151,22	270,31	836,38	2120,96	4038,48
1200	192,57	341,51	1021,97	2528,38	4852,41
Th	6,232	3,514	1,174	0,475	0,247
1/Th	0,160	0,285	0,852	2,107	4,044

Benchmarks

- BaseBatch avec k=30, modèle tiny

Model	tiny	base	small	medium	large
CPU rate			$\simeq 90\%$		
Memory (GB)	22.5	27.5	48	115	192
Throughput (Th)	39.8	30.1	6.5	4.5	2

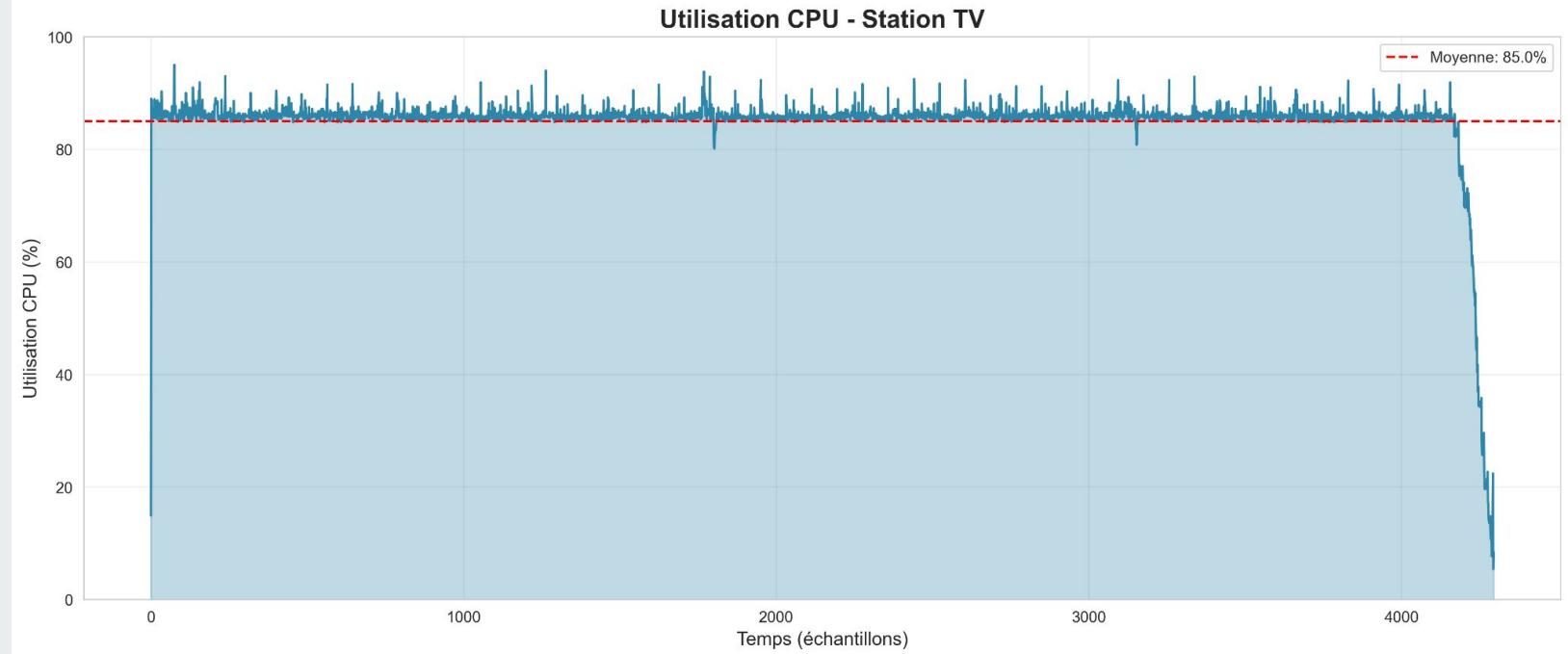
```
=====
RAPPORT QoS - STATION TV - TRANSCRIPTION AUDIO
=====

RÉSUMÉ DE LA SESSION
-----
Durée de la session: 13.16 heures
Nombre total de fichiers: 720
Fichiers réussis: 720
Fichiers échoués: 0
Taux de réussite: 100% 100%

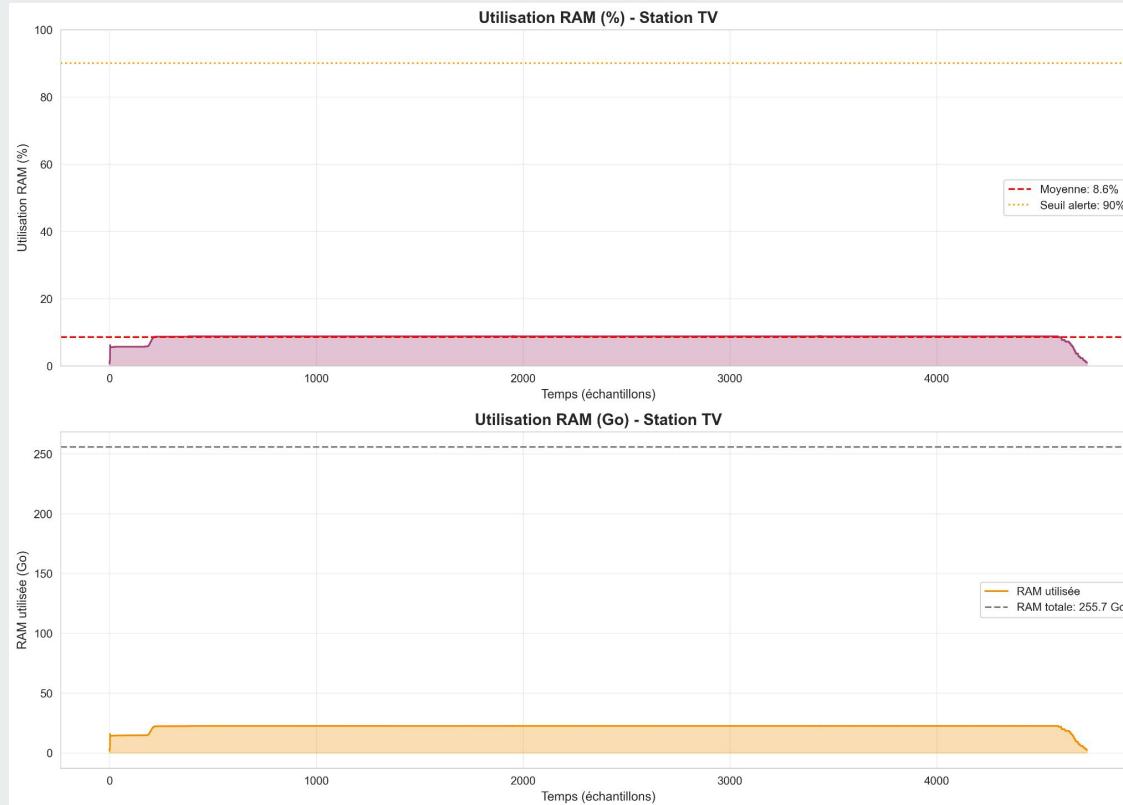
PERFORMANCE
-----
Durée audio totale traitée: 728 heures
Temps de traitement total: 13.16 heures
Throughput (Th): 1.01 (728h / 720 fichiers)
Temps moyen par fichier: 1947.44 secondes

OBJECTIFS QoS
-----
✓ Taux de réussite 100% >= 99%
✓ Traitement terminé avec succès
=====
```

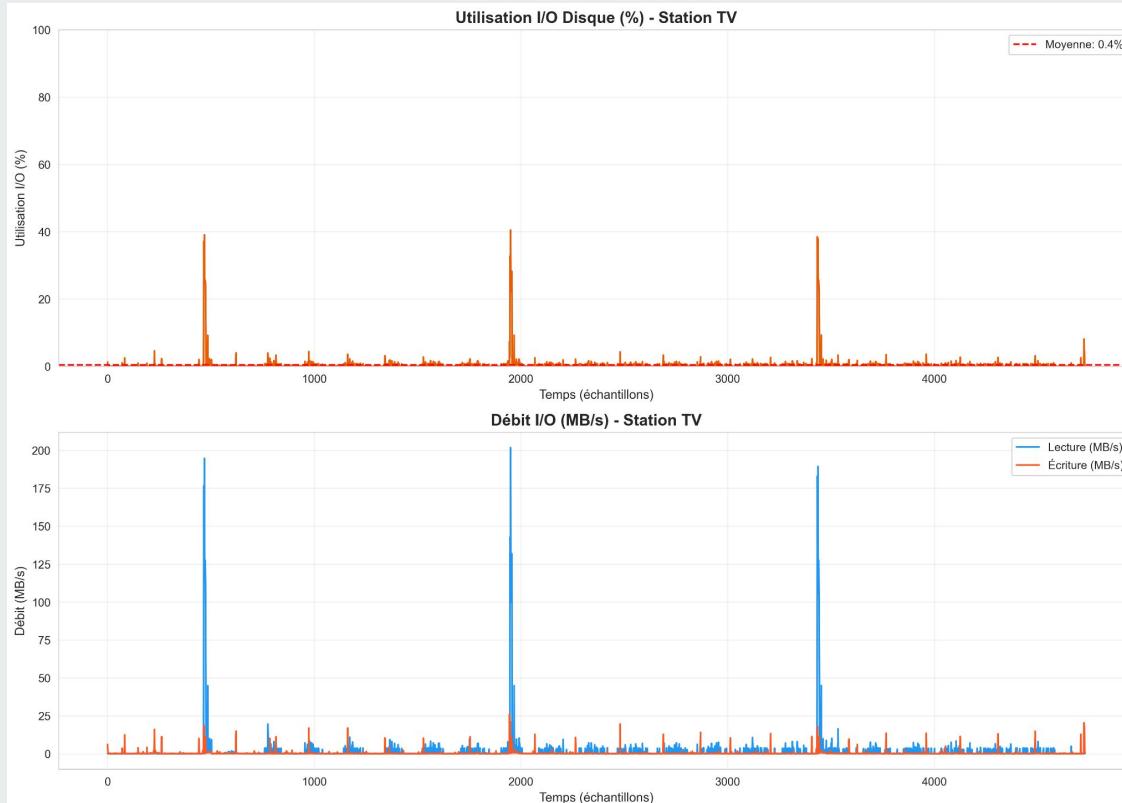
Benchmarks



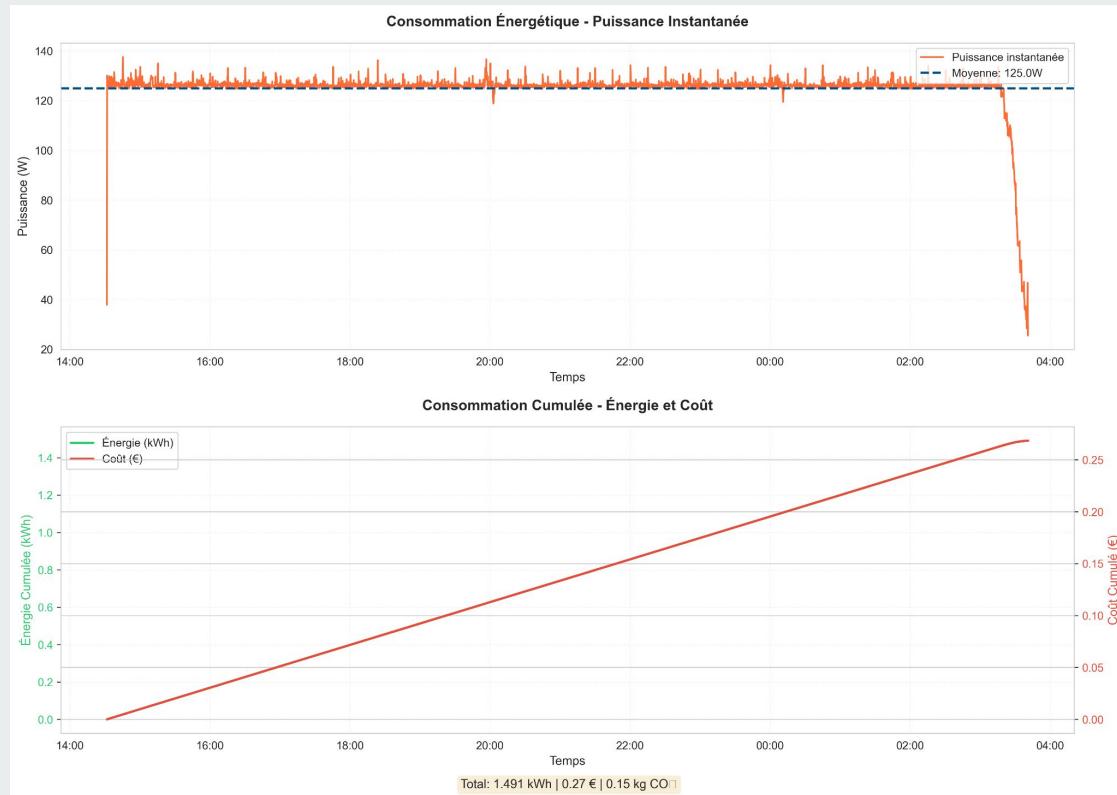
Benchmarks



Benchmarks



Benchmarks



Benchmarks

- BaseBatch avec k=30, modèle base

Model	tiny	base	small	medium	large
CPU rate			$\simeq 90\%$		
Memory (GB)	22.5	27.5	48	115	192
Throughput (Th)	39.8	30.1	6.5	4.5	2

Benchmarks

- BaseBatch avec k=30, modèle small

Model	tiny	base	small	medium	large
CPU rate	$\simeq 90\%$				
Memory (GB)	22.5	27.5	48	115	192
Throughput (Th)	39.8	30.1	6.5	4.5	2



Résultats

Résultats

- Throughput amélioré
- Th convaincant sur BaseMod
- Validation modèle tiny, base et small sur BaseBatch

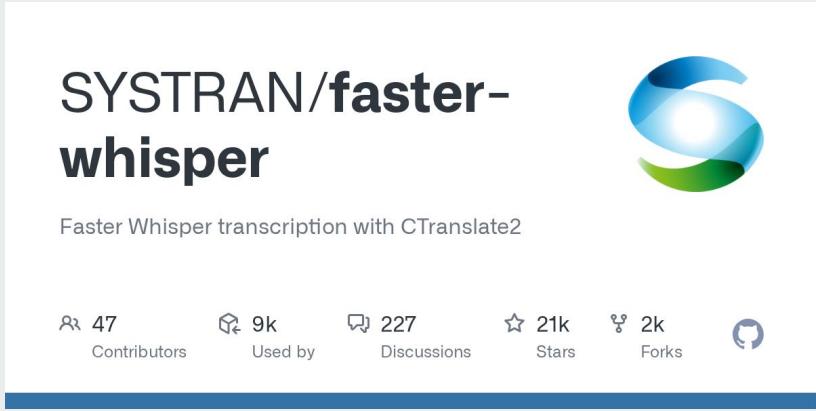


Conclusion et perspectives

Conclusion

- Pipeline exploitable
- Objectifs presque atteints : tests à faire sur medium et large
- Lancer des campagnes à grande échelle sur la base HelloWorld

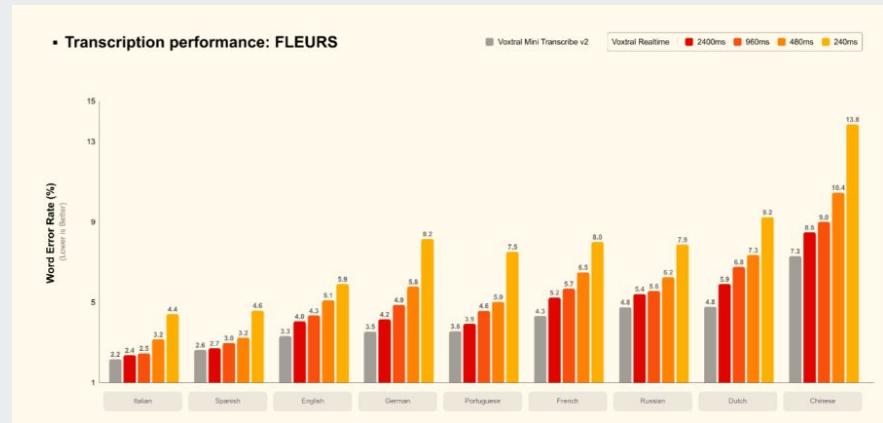
Perspectives



The GitHub repository page for SYSTRAN/faster-whisper shows the following details:

- Name:** SYSTRAN/faster-whisper
- Description:** Faster Whisper transcription with CTranslate2
- Contributors:** 47
- Used by:** 9k
- Discussions:** 227
- Stars:** 21k
- Forks:** 2k

faster-whisper



Voxtral Transcribe V2



Questions