# Rapport du test des outils ASR

## 1. Test du modèle MMS

Pour le test, j'ai essayé d'abord avec le corpus de Shanghaïen (la langue Wu) que j'ai trouvé la dernière fois dans une étude.

Le corpus du wu (audio_wuu) contient 50 enregistrements, après avoir utilisé le script de LID, 15 langues différentes ont été prédites. Dedans, la langue la plus fréquemment détectée est le mandarin (cmn), représentant 26 fichiers ; cela montre que le modèle confond souvent le wu avec le chinois standard. Les restes sont variés, comme vie, lao, mya, kor, hak, yue, jpn, etc. Mais je trouve que les prédictions restent géographiquement en Asie de l'est et de sud est.

Et puis j'essaie avec l'autre script de run_MMS.py, il y a une erreur :

ValueError: wuu does not exist.

J'ai changé le code en cmn et j'ai voulu essayer le résultat de la transcription. Aucune surprise, j'ai découvert qu'il reconnaissait les caractères chinois, mais ils étaient complètement dénués de sens. Il existe de nombreuses substitutions, plutôt dues à des remplacements phonétiquement proches. Cela montre aussi qu'il existe un grand écart entre le Wu et le mandarin.

J'ai donc choisi de tester le cantonais. Il utilise aussi des caractères chinois et est un dialecte chinois, mais il dispose de plus de ressources que le wu et le modèle le prend également en charge. L'objectif était d'observer comment le modèle se comporte sur le cantonais. Après l'avoir téléchargé depuis le site web Lingua Libre, j'ai aussi sélectionné les 50 fichiers audio, les ai d'abord convertis du format .ogg au format .wav, puis effectué le même test.

Pour le résultat de la prédiction, 12 langues ont été reconnues, la majorité étant correctement identifiée comme yue (cantonnais). Globalement, la détection du cantonais est plus cohérente que celle du wu, mais la diversité des prédictions indique encore des limitations importantes du modèle pour les dialectes.

En ce qui concerne la transcription, le modèle n'est pas très performant non plus. Il arrive que des fichiers texte soient vides. Je pense que cela est dû au fait que l'audio du corpus ne comporte qu'un seul mot, et non une phrase complète. Il est difficile d'évaluer la transcription en fonction du contexte ; il existe du coup des substitutions aussi. Cependant, en général, les résultats de transcription sont meilleurs qu'en langue wu.

Pour conclure, wu est un langage très peu doté ; MMS ne couvre pas wuu et ne peut que montrer les limitations. La performance du modèle en yue est un peu meilleure que wu, mais toujours pas très exacte.

## 2. Test du modèle Whisper et Qwen3-ASR

Comme le MMS ne couvre pas wuu, j'ai trouvé deux autres outils en ligne : Whisper et Qwen, et les testé par curiosité.

**Whisper (Openai)**
https://github.com/openai/whisper

Comme dans Whisper il n'y a pas de la langue Wu non plus, je le remplace par Chinese mandarin pour voir la distance entre deux langues en oraux, voici une capture d'écran de mon terminal :

```
(venv) sibel@sibelw:~/TestASR$ for f in *.wav; do
    whisper "$f" --model small --language Chinese --task transcribe
done
/home/sibel/venv/lib/python3.12/site-packages/whisper/transcribe.py:132: UserWarning: FP16 is not supported on CP
U; using FP32 instead
  warnings.warn("FP16 is not supported on CPU; using FP32 instead")
[00:00.000 --> 00:04.300] 沒有,放得在監獄有被女生發凌
/home/sibel/venv/lib/python3.12/site-packages/whisper/transcribe.py:132: UserWarning: FP16 is not supported on CP
U; using FP32 instead
  warnings.warn("FP16 is not supported on CPU; using FP32 instead")
[00:00.000 --> 00:02.000] 弄我的母人,对吧
/home/sibel/venv/lib/python3.12/site-packages/whisper/transcribe.py:132: UserWarning: FP16 is not supported on CP
U; using FP32 instead
  warnings.warn("FP16 is not supported on CPU; using FP32 instead")
[00:00.000 --> 00:02.440] 太過分了 可能會知道
/home/sibel/venv/lib/python3.12/site-packages/whisper/transcribe.py:132: UserWarning: FP16 is not supported on CP
U; using FP32 instead
  warnings.warn("FP16 is not supported on CPU; using FP32 instead")
[00:00.000 --> 00:02.000] 呵呵
[00:02.000 --> 00:04.000] 得为至终至少二斤
/home/sibel/venv/lib/python3.12/site-packages/whisper/transcribe.py:132: UserWarning: FP16 is not supported on CP
U; using FP32 instead
  warnings.warn("FP16 is not supported on CPU; using FP32 instead")
[00:00.000 --> 00:02.000] 坤果是都全顏碼子
[00:02.000 --> 00:04.000] 例子由杰米的功术
[00:04.000 --> 00:06.000] 明白
```

Voici un tableau de comparaison :

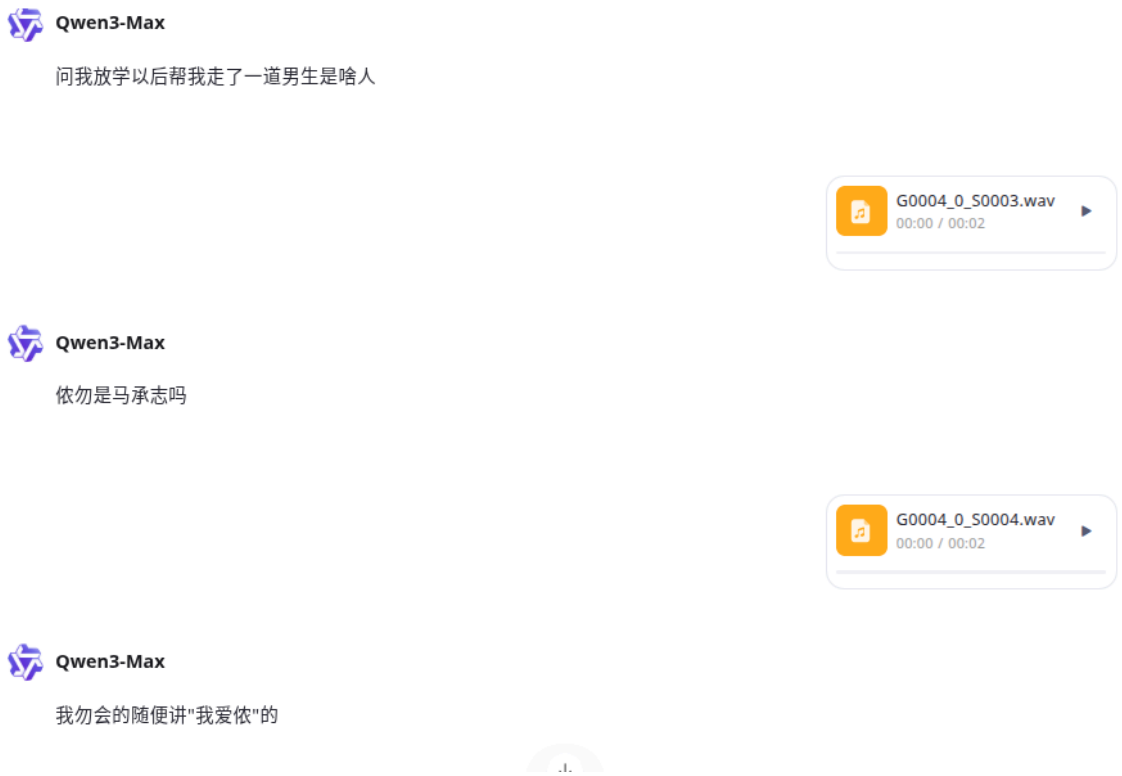
| | Transcription manuelle | Transcription Whisper | Traduction en mandarin |
|---|---|---|---|
| Phrase 1 | 问吾 放学以后和吾走辣一道呃男生是撒拧 | 沒有,放得在監獄有被女生發凌 | 问我 放学以后和我走的那些男生是谁 |
| Phrase 2 | 侬伐是马成对伐？ | 弄我的母人,对吧 | 你不是马成对吧 |
| Phrase 3 | 吾伐会的随便讲吾爱侬呃 | 太過分了 可能會知道 | 我不会随便说我爱你 |
| Phrase 4 | 呵呵，德汇至尊至尚呃紧 | 呵呵<br>得为至终至少二斤 | 呵呵 德汇至尊至尚的警 |
| Phrase 5 | 困觉前头吃眼物事，李子有减肥功效明白 | 坤果是都全顏碼子<br>例子由杰米的功术<br>明白 | 睡前吃点东西 李子有减肥功效明白 |

On trouve que le modèle produit du chinois standard relativement fluent mais dévie fortement du sens de l'énoncé Wu. Il existe aussi de nombreuses substitutions avec des mots inexistants dans l'audio. La récupération de mots clés est également très limitée, seuls quelques mots comme "呵呵", "明白" sont corrects, mais probablement en raison de leur proximité phonétique avec le mandarin.

En gros, Whisper a une compréhension globale faible de Wu.

**Qwen3 (Alibaba)**
https://github.com/QwenLM
https://chat.qwen.ai/c/guest

Les codes sont un peu compliqués du coup j'utilise directement le website de Qwen3 pour transcrire, voici une capture d'écran de la transcription :




Aussi un tableau de comparaison :

| | Transcription manuelle | Transcription Qwen3 | Traduction en mandarin |
|---|---|---|---|
| Phrase 1 | 问吾 放学以后和吾走辣一道呃男生是撒拧 | 问我放学以后帮我走了一道男生是啥人 | 问我 放学以后和我走的那些男生是谁 |
| Phrase 2 | 侬伐是马成对伐？ | 侬勿是马承志吗 | 你不是马成对吧 |
| Phrase 3 | 吾伐会的随便讲吾爱侬呃 | 我勿会的随便讲"我爱侬"的 | 我不会随便说我爱你 |
| Phrase 4 | 呵呵，德汇至尊至尚呃紧 | 呵呵迭个会自作自受的劲 | 呵呵 德汇至尊至尚的警 |
| Phrase 5 | 困觉前头吃眼物事，李子有减肥功效明白 | 困觉前头吃眼么子例如有减肥的功效明白 | 睡前吃点东西 李子有减肥功效明白 |

| | | | |
|---|---|---|---|
| | | | |

Le modèle Qwen3-ASR reconnaît bien le shanghaïen. Les phrases sont fluides et le sens global est correct, avec seulement quelques différences d'écriture comme *"吾/我"* ou *"勿/伐"* (mais les sens sont pareils).
La transcription reste claire et naturelle, montrant une bonne adaptation du modèle au Wu.

En résumé, ces résultats illustrent la différence de performance entre une langue peu dotée et une langue mieux dotée.