

[illegible]

2) Entraînement du modèle :

```
(dl_audio_env) C:\Users\Boumahdi\Documents\hayat\S2\ml\projet final\transcription>python train.py
Dataset chargé depuis les fichiers locaux
{'client_id': '89fa8146b2c07e3e2fe2bbf3852ed8b22a625e383df7eb7225688c8ca9ecc3a151cdb4447121c0a927b1e01eb6471c855f06accb80f1ec0fac9e257c03286895', 'path': 'common_voice_fr_21982258.mp3', 'sentence': 'Firefox', 'up_votes': 2, 'down_votes': 0, 'age': 'teens', 'gender': 'male', 'accent': 'france', 'locale': 'fr', 'segment': 'Benchmark', 'audio': {'path': 'C:\\Users\\Boumahdi\\Documents\\hayat\\S2\\ml\\projet final\\transcription\\data\\cv-corpus-7.0-singleword\\fr\\clips\\common_voice_fr_21982258.mp3', 'array': array([ 0.0000000e+00,  0.0000000e+00,  0.0000000e+00, ..., -2.25731310e-06, -7.40760152e-06, -1.41151959e-05], shape=(48384,)), 'sampling_rate': 16000}}
Epoch 1 - Loss: 4.611921991384437
Epoch 2 - Loss: 0.4678938132017217
Epoch 3 - Loss: 0.3893697382081007
Epoch 4 - Loss: 0.34572145655737446
Epoch 5 - Loss: 0.3326355184494732
```

3) Evaluation du modèle :

```
(dl_audio_env) C:\Users\Boumahdi\Documents\hayat\S2\ml\projet final\transcription>python run_model.py
Dataset chargé depuis les fichiers locaux
{'client_id': '89fa8146b2c07e3e2fe2bbf3852ed8b22a625e383df7eb7225688c8ca9ecc3a151cdb4447121c0a927b1e01eb6471c855f06accb80f1ec0fac9e257c03286895', 'path': 'common_voice_fr_21982258.mp3', 'sentence': 'Firefox', 'up_votes': 2, 'down_votes': 0, 'age': 'teens', 'gender': 'male', 'accent': 'france', 'locale': 'fr', 'segment': 'Benchmark', 'audio': {'path': 'C:\\Users\\Boumahdi\\Documents\\hayat\\S2\\ml\\projet final\\transcription\\data\\cv-corpus-7.0-singleword\\fr\\clips\\common_voice_fr_21982258.mp3', 'array': array([ 0.0000000e+00,  0.0000000e+00,  0.0000000e+00, ..., -2.25731310e-06, -7.40760152e-06, -1.41151959e-05], shape=(48384,)), 'sampling_rate': 16000}}
Map: 100% | 8648/8648 [09:13<00:00, 15.61 examples/s]
Map: 100% | 20017/20017 [21:25<00:00, 15.57 examples/s]
Map: 100% | 5685/5685 [1:03:06<00:00, 1.50 examples/s]
Validation Loss: 0.9237437500790521
```

III) Commande vocale :

1) Installation des libraires nécessaires :

```
(dl_audio_env) C:\WINDOWS\system32>pip install torch torchaudio librosa matplotlib numpy pandas scikit-learn streamlit gradio
Collecting torch
  Downloading torch-2.7.0-cp310-cp310-win_amd64.whl (212.5 MB)
  ----- 212.5/212.5 MB 1.5 MB/s eta 0:00:00
Collecting torchaudio
  Downloading torchaudio-2.7.0-cp310-cp310-win_amd64.whl (2.5 MB)
  ----- 2.5/2.5 MB 2.0 MB/s eta 0:00:00
Collecting librosa
  Downloading librosa-0.11.0-py3-none-any.whl (260 kB)
  ----- 260.7/260.7 kB 1.6 MB/s eta 0:00:00
Collecting matplotlib
  Downloading matplotlib-3.10.3-cp310-cp310-win_amd64.whl (8.1 MB)
  ----- 8.1/8.1 MB 2.5 MB/s eta 0:00:00
Collecting numpy
  Using cached numpy-2.2.5-cp310-cp310-win_amd64.whl (12.9 MB)
Collecting pandas
  Using cached pandas-2.2.3-cp310-cp310-win_amd64.whl (11.6 MB)
Collecting scikit-learn
  Downloading scikit_learn-1.6.1-cp310-cp310-win_amd64.whl (11.1 MB)
  ----- 11.1/11.1 MB 853.2 kB/s eta 0:00:00
Collecting streamlit
  Downloading streamlit-1.45.1-py3-none-any.whl (9.9 MB)
  ----- 9.9/9.9 MB 1.7 MB/s eta 0:00:00
Collecting gradio
  Using cached gradio-5.29.1-py3-none-any.whl (54.1 MB)
Collecting typing-extensions>=4.10.0
  Using cached typing_extensions-4.13.2-py3-none-any.whl (45 kB)
Collecting networkx
  Downloading networkx-3.4.2-py3-none-any.whl (1.7 MB)
  ----- 1.7/1.7 MB 2.0 MB/s eta 0:00:00
Collecting fsspec
  Using cached fsspec-2025.3.2-py3-none-any.whl (194 kB)
Collecting sympy>=1.13.3
  Downloading sympy-1.14.0-py3-none-any.whl (6.3 MB)
  ----- 6.3/6.3 MB 2.3 MB/s eta 0:00:00
```

```
(dl_audio_env) C:\Users\Boumahdi\Documents\hayat\S2\ml\projet final>pip install seaborn
Collecting seaborn
  Downloading seaborn-0.13.2-py3-none-any.whl (294 kB)
    ----- 294.9/294.9 kB 1.7 MB/s eta 0:00:00
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in c:\windows\system32\dl_audio_env\lib\site-packages (from seaborn) (3.10.3)
Requirement already satisfied: pandas>=1.2 in c:\windows\system32\dl_audio_env\lib\site-packages (from seaborn) (2.2.3)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in c:\windows\system32\dl_audio_env\lib\site-packages (from seaborn) (2.2.5)
Requirement already satisfied: pyparsing>=2.3.1 in c:\windows\system32\dl_audio_env\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.2.3)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\windows\system32\dl_audio_env\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.8)
Requirement already satisfied: pillow>=8 in c:\windows\system32\dl_audio_env\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (11.2.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\windows\system32\dl_audio_env\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.58.0)
Requirement already satisfied: python-dateutil>=2.7 in c:\windows\system32\dl_audio_env\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (2.9.0.post0)
Requirement already satisfied: packaging>=20.0 in c:\windows\system32\dl_audio_env\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (24.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\windows\system32\dl_audio_env\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.3.2)
Requirement already satisfied: cycler>=0.10 in c:\windows\system32\dl_audio_env\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)
Requirement already satisfied: pytz>=2020.1 in c:\windows\system32\dl_audio_env\lib\site-packages (from pandas>=1.2->seaborn) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in c:\windows\system32\dl_audio_env\lib\site-packages (from pandas>=1.2->seaborn) (2025.2)
Requirement already satisfied: six>=1.5 in c:\windows\system32\dl_audio_env\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.17.0)
Installing collected packages: seaborn
Successfully installed seaborn-0.13.2
```

2) Téléchargement des données :

```
(dl_audio_env) C:\Users\Boumahdi\Documents\hayat\S2\ml\projet final>python download_dataset.py
Téléchargement en cours...
Téléchargement terminé.
Extraction en cours...
Extraction terminée.
```

3) L'entraînement du modèle :

```
(dl_audio_env) C:\Users\Boumahdi\Documents\hayat\S2\ml\projet final>python train.py
Nombre batches entraînement : 964
Nombre batches test : 241
Appareil utilisé : cpu
Début époque 1
  Batch 1/964
  Batch 2/964
  Batch 3/964
  Batch 4/964
  Batch 5/964
  Batch 6/964
  Batch 7/964
  Batch 8/964
  Batch 9/964
  Batch 10/964
  Batch 11/964
  Batch 12/964
  Batch 13/964
  Batch 14/964
  Batch 15/964
  Batch 16/964
  Batch 17/964
  Batch 18/964
  Batch 19/964
  Batch 20/964
  Batch 21/964
  Batch 22/964
  Batch 23/964
  Batch 24/964
```

```
  Batch 962/964
  Batch 963/964
  Batch 964/964
Époque 5 - Perte moyenne : 0.3134
[ ] Modèle sauvegardé : simple_cnn_speech_commands.pth
```

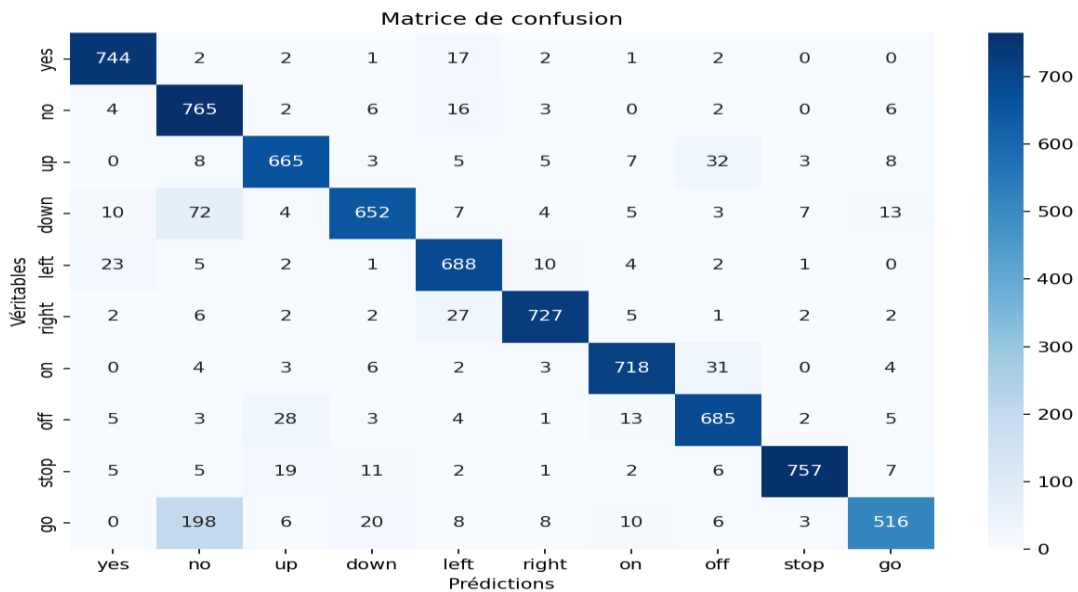
4) L'évaluation du modèle :

```
(dl_audio_env) C:\Users\Boumahdi\Documents\hayat\S2\ml\projet final>python evaluate.py
Rapport de classification :
precision    recall  f1-score   support

yes         0.94     0.96     0.95       771
no          0.72     0.95     0.82       804
up          0.91     0.90     0.91       736
down        0.92     0.84     0.88       777
left        0.89     0.93     0.91       736
right       0.95     0.94     0.94       776
on          0.94     0.93     0.93       771
off         0.89     0.91     0.90       749
stop        0.98     0.93     0.95       815
go          0.92     0.67     0.77       775

accuracy    0.90
macro avg   0.90
weighted avg 0.90

Matrice de confusion :
[[744  2  2  1 17  2  1  2  0  0]
 [ 4 765  2  6 16  3  0  2  0  6]
 [ 0  8 665  3  5  5  7 32  3  8]
 [10 72  4 652  7  4  5  3  7 13]
 [23  5  2  1 688 10  4  2  1  0]
 [ 2  6  2  2 27 727  5  1  2  2]
 [ 0  4  3  6  2  3 718 31  0  4]
 [ 5  3 28  3  4  1 13 685  2  5]
 [ 5  5 19 11  2  1  2  6 757  7]
 [ 0 198  6 20  8  8 10  6  3 516]]
```



IV) Création de l'interface :

```
C:\WINDOWS\system32>dl_audio_env\Scripts\activate

(dl_audio_env) C:\WINDOWS\system32>cd C:\Users\Boumahdi\Documents\hayat\S2\ml\projet final

(dl_audio_env) C:\Users\Boumahdi\Documents\hayat\S2\ml\projet final>python app.py
* Running on local URL: http://127.0.0.1:7860
* Running on public URL: https://d99a71102d27cdd2db.gradio.live

This share link expires in 1 week. For free permanent hosting and GPU upgrades, run `gradio deploy` from the terminal in
the working directory to deploy to Hugging Face Spaces (https://huggingface.co/spaces)
```

127.0.0.1:7860

Détection vocale et transcription

Choisissez une tâche, puis importez un fichier audio pour obtenir soit la transcription, soit la commande détectée.

Fichier audio (wav/mp3/flac/ogg)

↑

Déposez l'audio ici

-OU-

Cliquez pour télécharger un fichier

sortir

Marqueur

Choix de la tâche

☐ Transcription ☐ Commande vocale

Effacer Envoyer

Transcription & Détection Commande Vocale

Uploader un fichier audio et choisissez la tâche.

Uploader un fichier audio



Choisir la tâche

☒ Transcription vocale ☐ Détection commande vocale

Effacer Envoyer

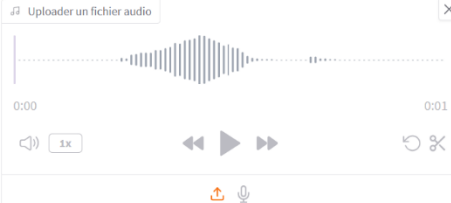
Résultat

Transcription : Je voudrais oublier le temps pour un souvenir pour un instant une parenthèse après la course et partir où mon cœur me pousse Je voudrais retrouver mes traces où est ma vie où est ma place et garder leur demande passé mon chaud dans mon jardin secret Je voudrais passer l'océan croisé le vol de Gaulleant passer à tout ce que j'ai vu ou bien aller vers l'inconnu Je voudrais décrocher la lune, je voudrais même sauver l'entrée mais ça vende tout Je voudrais parler à mon veille parler à mon veille Je voudrais choisir un bateau pas le plus grand et le plus beau Je leur entierai des images et des parfums de mes voyage Je voudrais frais-l'épau ma soeur trouver au cru de ma mémoire des voies de ceux qui montent à prix qu'il n'y a pas de rêve interdit Je voudrais trouver l'épauille du tableau que j'ai dans l'oucre de se décoller plus où je vous vois et m'en m'assure je voudrais décrocher la lune, je voudrais même sauver l'entrée mais ça vende tout Je voudrais parler à mon veille parler à mon veille Je voudrais courbier le temps pour un souviure pour un instant qui me parenthèse après la course et parti où mon coeur me pousse je voudrais retrouver mes traces où est ma vie ou est ma classe et gare-ville l'homme m'en passer au jour dans mon jardin se fait je

Transcription & Détection Commande Vocale

Uploader un fichier audio et choisissez la tâche.

Uploader un fichier audio



Choisir la tâche

☐ Transcription vocale ☒ Détection commande vocale

Effacer Envoyer

Résultat

Commande détectée : left

Share via Link