

Projet 15 : Analyser en direct les flux de chaînes TV d'information en continu

Porteur de projet : Baptiste Queuche

Présenté par Damien Conceicao, Maria Kirbasova, Marisol de Nazelle, Garance Raynaud, et Manon Remot

Contexte du projet

La **diarisation vocale** consiste à identifier “**qui parle** et **quand**” dans un enregistrement audio.

Objectifs :

- Construire un workflow permettant de décompter le **temps de parole réel** grâce à une **identification vocale**.
- Comparer avec les temps déclarés à l'ARCOM



InfoVox Tracker



Données

■ Corpus sonore

Intégralité de la diffusion
de Cnews de 7h à 00h, en
fichiers audio mp3 d'1h

■ Database locuteurs

Fichiers csv correspondant aux
bandeaux de Cnews pour savoir
qui parle (id, nom, prénom, temps
de parole)



Fichiers csv correspondant
uniquement aux émissions
(sans les pubs)



■ Database locuteurs

Speaker	Debut	Fin	Durée	Images	Direct	Programme	ID
ADRIEN FONTENEAU @AdriFnin	2025-12-09 00:06:09	2025-12-09 00:06:09	0 days 00:00:05	1	0	Absent	124757
SEBASTIEN LECORNU PREMIER MINISTRE	2025-12-09 00:08:11	2025-12-09 00:08:24	0 days 00:00:18	3	0	L'ÉDITION DE LA NUIT	1834
ANNIE GENEVARD MINISTRE DE	2025-12-09 00:09:57	2025-12-09 00:10:24	0 days 00:00:32	6	0	L'ÉDITION DE LA NUIT	2538
TANGUY HAMON JOURNALISTE	2025-12-09 00:11:17	2025-12-09 00:12:19	0 days 00:00:47	9	0	L'ÉDITION DE LA NUIT	35
JULIETTE SADAT JOURNALISTE	2025-12-09 00:12:54	2025-12-09 00:13:51	0 days 00:01:02	13	0	L'ÉDITION DE LA NUIT	3507
OLIVIER MADINIER EN DUPLEX DE	2025-12-09 00:16:02	2025-12-09 00:16:47	0 days 00:00:50	10	0	L'ÉDITION DE LA NUIT	367
FRANCK ROBINE PRÉFET DE LA	2025-12-09 00:17:54	2025-12-09 00:18:18	0 days 00:00:29	6	0	L'ÉDITION DE LA NUIT	1755
EMMANUEL MACRON PRÉSIDENT	2025-12-09 00:19:13	2025-12-09 00:19:48	0 days 00:00:40	8	0	L'EDITION DE LA NUIT	104
ARTHUR VERMEEREN MILIEU DE TERRAIN	2025-12-09 00:23:24	2025-12-09 00:23:24	0 days 00:00:05	1	0	L'ÉDITION DE LA NUIT	133689
CPL CIMAL CIM EH) ETHAD ETHD CS	2025-12-09 00:24:16	2025-12-09 00:24:16	0 days 00:00:05	1	0	Absent	133699

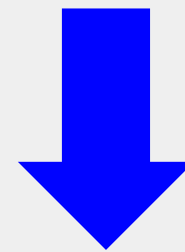
Nettoyage des données

suppression pubs, séquençage 5min



Création vérité terrain

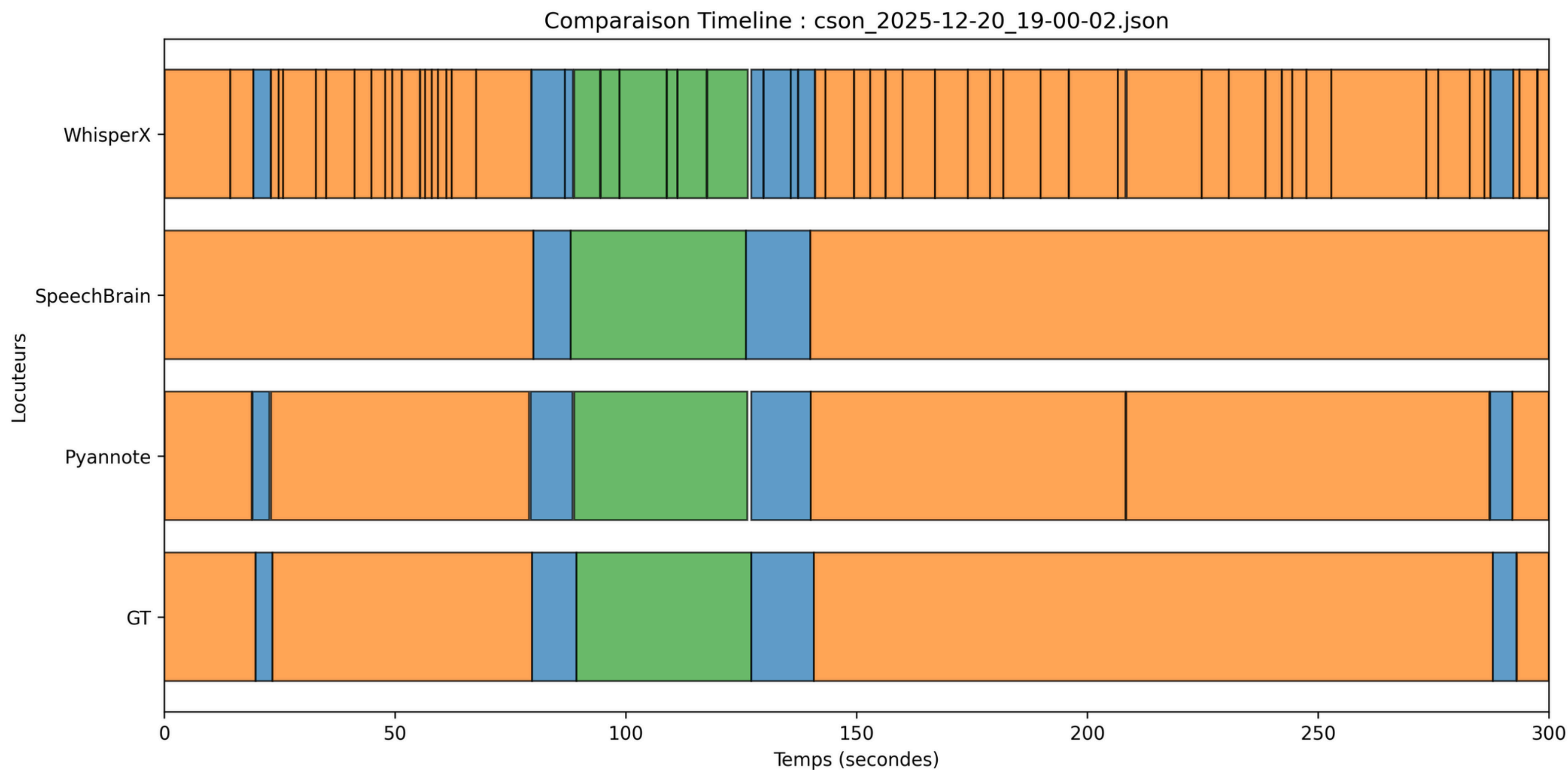
annotations manuelles 12 fichiers
audios



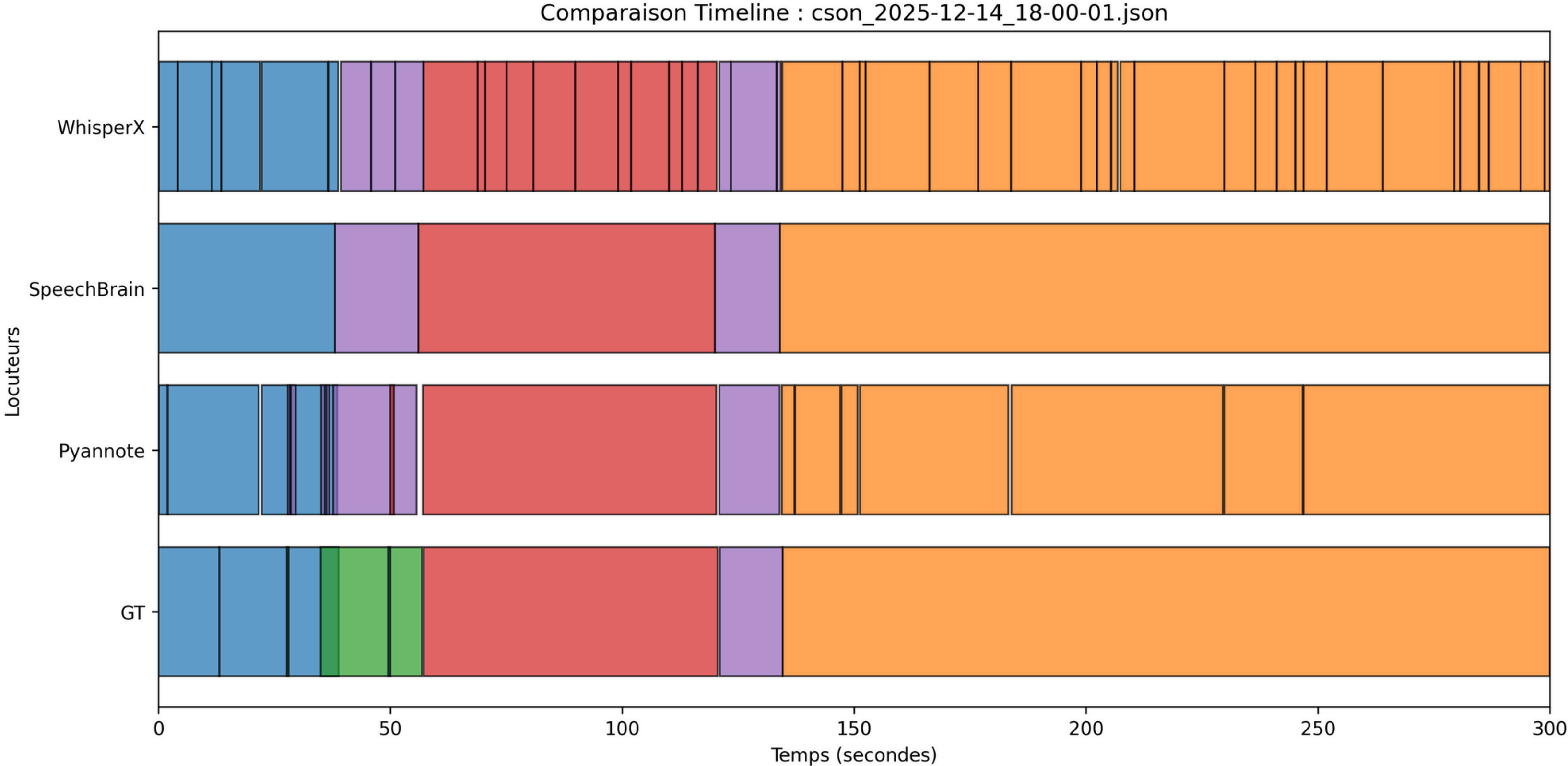
Analyse 3 modèles

Pyannote, WhisperX, SpeechBrain

Comparaison des modèles



Comparaison des modèles



Passage à l'échelle

- Application du **modèle Pyannote** sur l'ensemble des données (émissions d'1h sans publicité)
- Préalablement séquencées en blocs de 5 minutes
- Identification des speakers / personnes
- Développement d'un code de **rapprochement voix ↔ personne** (id présents dans les csv)

Résultat :

Attribution automatique des interventions vocales à des personnes identifiées



<u>start</u>	<u>end</u>	<u>duration</u>	<u>Speaker_ID</u>	<u>unique_id</u>	<u>embeddings</u>
2025-12-08 11:00:00.030	2025-12-08 11:00:39.570	39.54	Unknown	3333836ebc	0.19372767 -0.26411888 0.12119127 0.10753789 0.03428236 0.16268033 -0.1677313 -0.12076437 0.40523747 0.14595152]
2025-12-08 11:00:40.500	2025-12-08 11:01:08.760	28.27	Unknown	3333836ebc	0.19372767 -0.26411888 0.12119127 0.10753789 0.03428236 0.16268033 -0.1677313 -0.12076437 0.40523747 0.14595152]
2025-12-08 11:01:09.470	2025-12-08 11:01:29.690	20.22	Unknown	3333836ebc	0.19372767 -0.26411888 0.12119127 0.10753789 0.03428236 0.16268033 -0.1677313 -0.12076437 0.40523747 0.14595152]
2025-12-08 11:01:29.940	2025-12-08 11:01:54.560	24.62	Unknown	a08665f05e	-0.3613509 -0.19229297 -0.22284546 -0.04171446 0.21419534 0.00882743 0.09770755 -0.06050294 -0.24882428 0.22831163]
2025-12-08 11:01:55.470	2025-12-08 11:02:02.140	6.67	1755	8769882bc5	-2.00127661e-01 -2.92866766e-01 1.85348228e-01 -2.85752621e-02 -9.21256021e-02 8.79466459e-02 1.93583250e-01 -3.70809846e-02]
2025-12-08 11:02:28.430	2025-12-08 11:02:30.990	2.56	1755	8769882bc5	-2.00127661e-01 -2.92866766e-01 1.85348228e-01 -2.85752621e-02 -9.21256021e-02 8.79466459e-02 1.93583250e-01 -3.70809846e-02]
2025-12-08 11:02:31.530	2025-12-08 11:02:52.510	20.98	Unknown	a08665f05e	-0.3613509 -0.19229297 -0.22284546 -0.04171446 0.21419534 0.00882743 0.09770755 -0.06050294 -0.24882428 0.22831163]
2025-12-08 11:02:53.910	2025-12-08 11:03:00.810	6.9	3987	a117d38445	-6.01330698e-02 -1.87368989e-01 3.20255041e-01 -8.75020176e-02 -3.82177830e-02 1.04738854e-01 -4.01733160e-01 5.73245659e-02]
2025-12-08 11:03:00.460	2025-12-08 11:03:00.480	0.02	84307	095dd7bd95	-0.1632711 0.15493746 -0.16153327 0.13430783 0.30977952 -0.14176777 -0.15073527 0.12554336 0.13491867 -0.01768493]
2025-12-08 11:03:00.480	2025-12-08 11:03:00.910	0.44	1755	8769882bc5	-2.00127661e-01 -2.92866766e-01 1.85348228e-01 -2.85752621e-02 -9.21256021e-02 8.79466459e-02 1.93583250e-01 -3.70809846e-02]
2025-12-08 11:03:00.910	2025-12-08 11:03:00.930	0.02	3987	a117d38445	-6.01330698e-02 -1.87368989e-01 3.20255041e-01 -8.75020176e-02 -3.82177830e-02 1.04738854e-01 -4.01733160e-01 5.73245659e-02]
2025-12-08 11:03:00.930	2025-12-08 11:03:01.030	0.1	1755	8769882bc5	-2.00127661e-01 -2.92866766e-01 1.85348228e-01 -2.85752621e-02 -9.21256021e-02 8.79466459e-02 1.93583250e-01 -3.70809846e-02]
2025-12-08 11:03:01.030	2025-12-08 11:03:05.990	4.96	3987	a117d38445	-6.01330698e-02 -1.87368989e-01 3.20255041e-01 -8.75020176e-02 -3.82177830e-02 1.04738854e-01 -4.01733160e-01 5.73245659e-02]
2025-12-08 11:03:01.230	2025-12-08 11:03:02.890	1.65	1755	8769882bc5	-2.00127661e-01 -2.92866766e-01 1.85348228e-01 -2.85752621e-02 -9.21256021e-02 8.79466459e-02 1.93583250e-01 -3.70809846e-02]
2025-12-08 11:03:02.890	2025-12-08 11:03:02.910	0.02	84307	095dd7bd95	-0.1632711 0.15493746 -0.16153327 0.13430783 0.30977952 -0.14176777 -0.15073527 0.12554336 0.13491867 -0.01768493]
2025-12-08 11:03:06.400	2025-12-08 11:03:25.970	19.58	3987	a117d38445	-6.01330698e-02 -1.87368989e-01 3.20255041e-01 -8.75020176e-02 -3.82177830e-02 1.04738854e-01 -4.01733160e-01 5.73245659e-02]

Document (presque) final d'identification des tours de parole

Tentative de regroupement des speakers inconnus

INPUT

Dataset : ($x_1 \dots x_N$)
 K_{\min}
 K_{\max}
Centroids connus

INITIALISATION CONTRAINTE

Fixer les K_{\min}
centroids connus

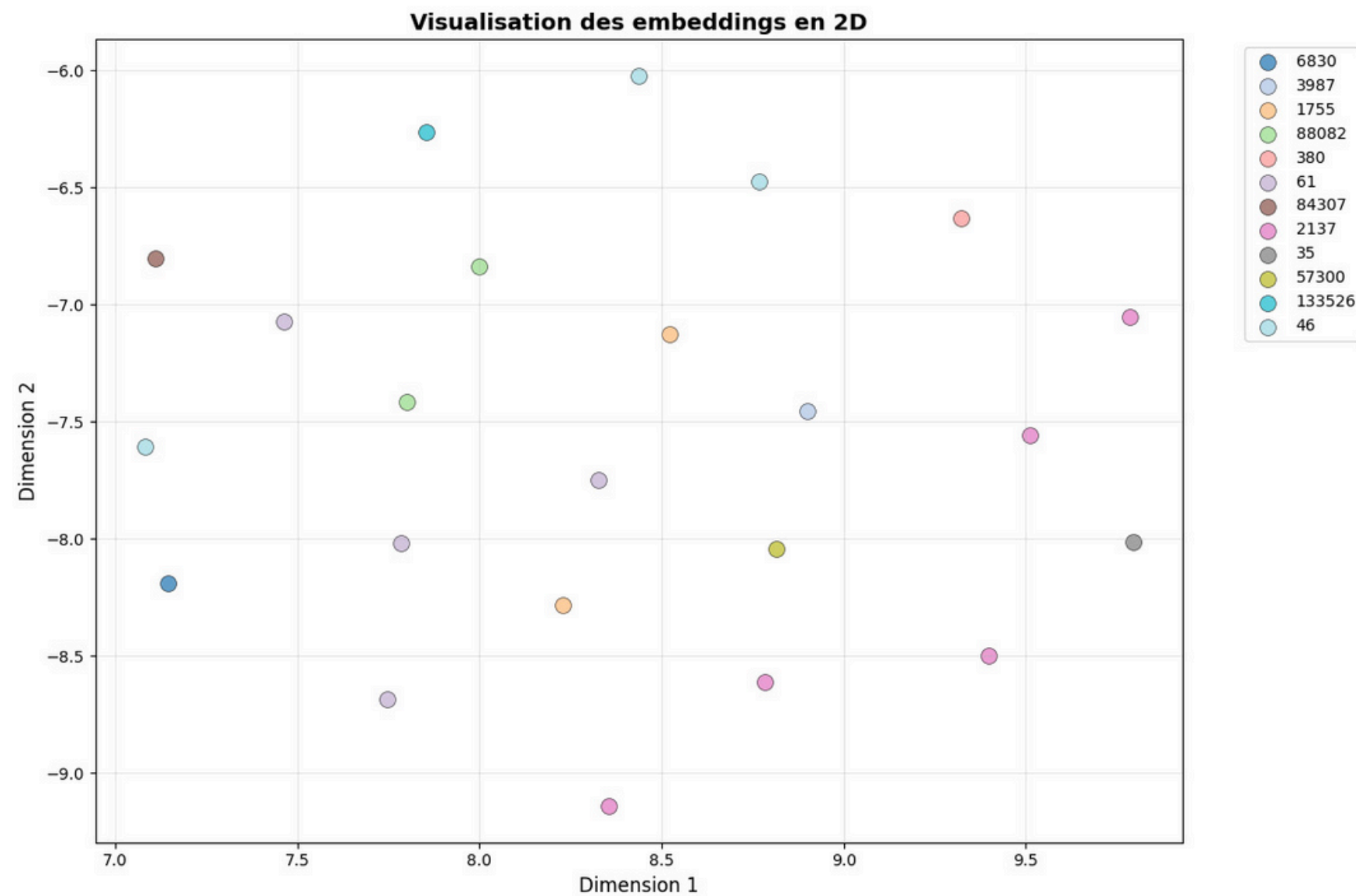
CRÉATION DES CENTROIDS SUPPLÉMENTAIRES (si $K > K_{\min}$)

Identifier le cluster à variance
maximale
Perturber ce centroid pour créer μ_k

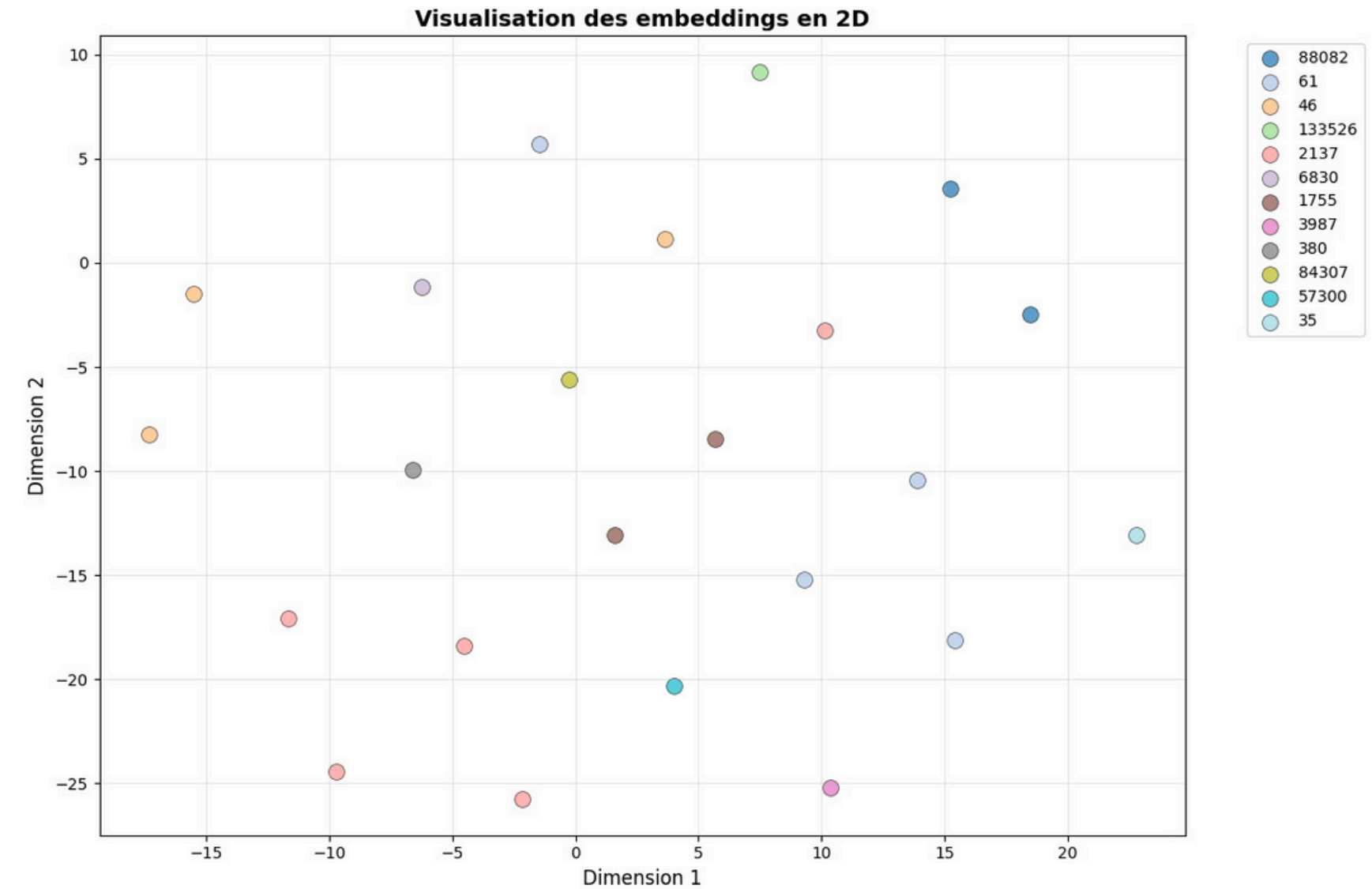
BOUCLE PRINCIPALE ($K = K_{\min} \rightarrow K_{\max}$)

Répéter jusqu'à convergence

Résultats 1/2

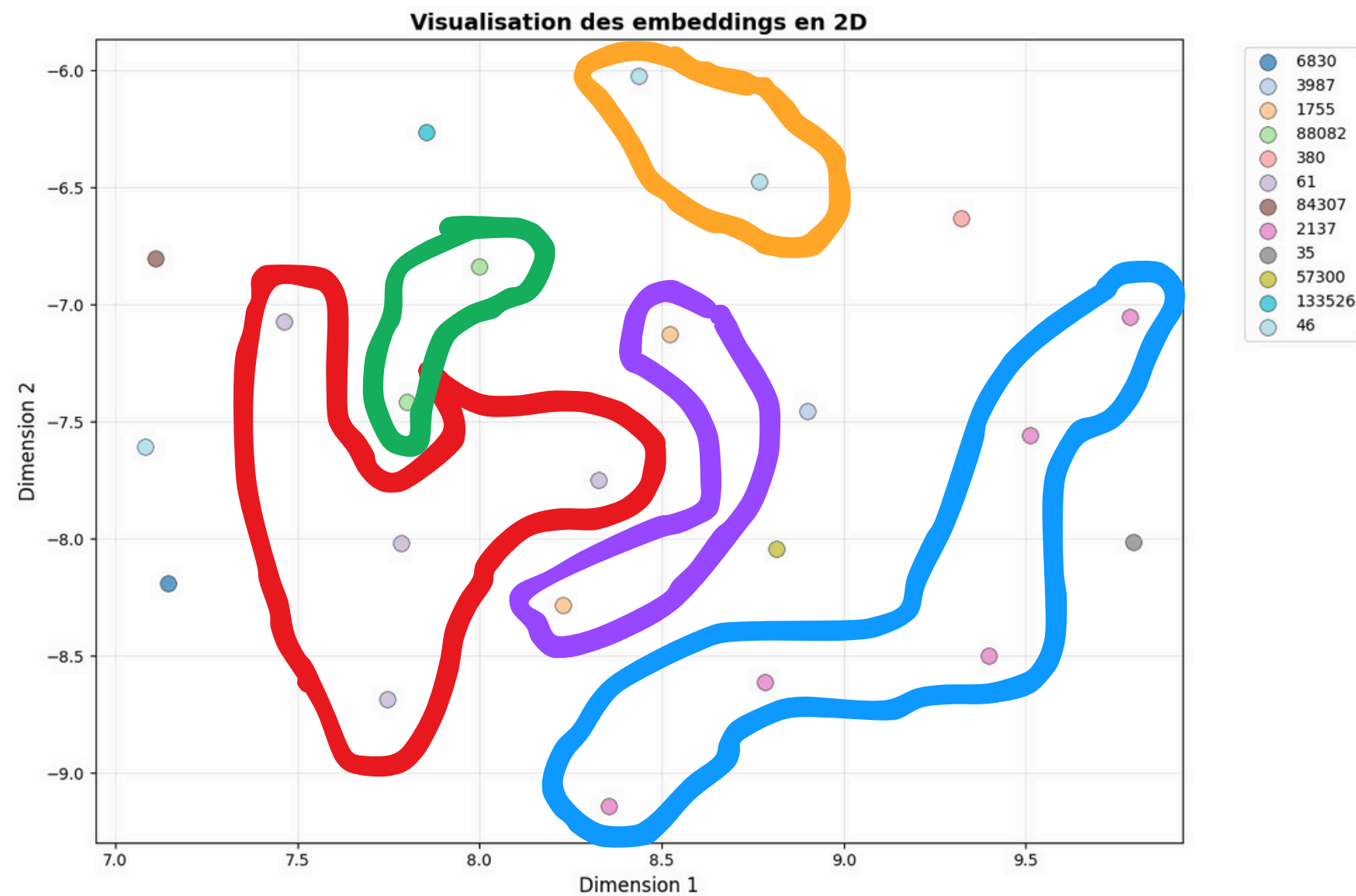


projection 2D des embeddings avec UMAP



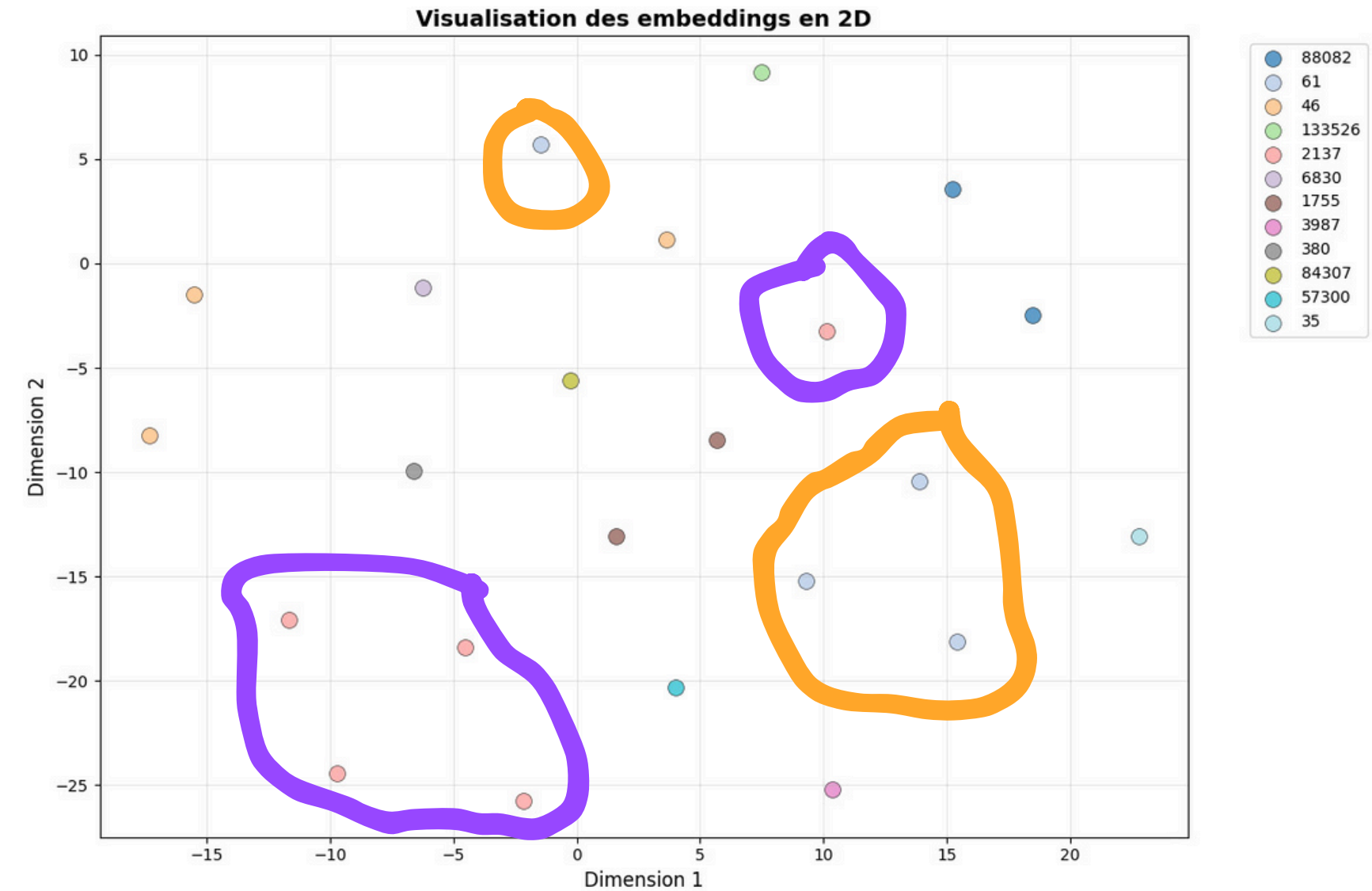
La même chose avec t-SNE

Résultats 1/2



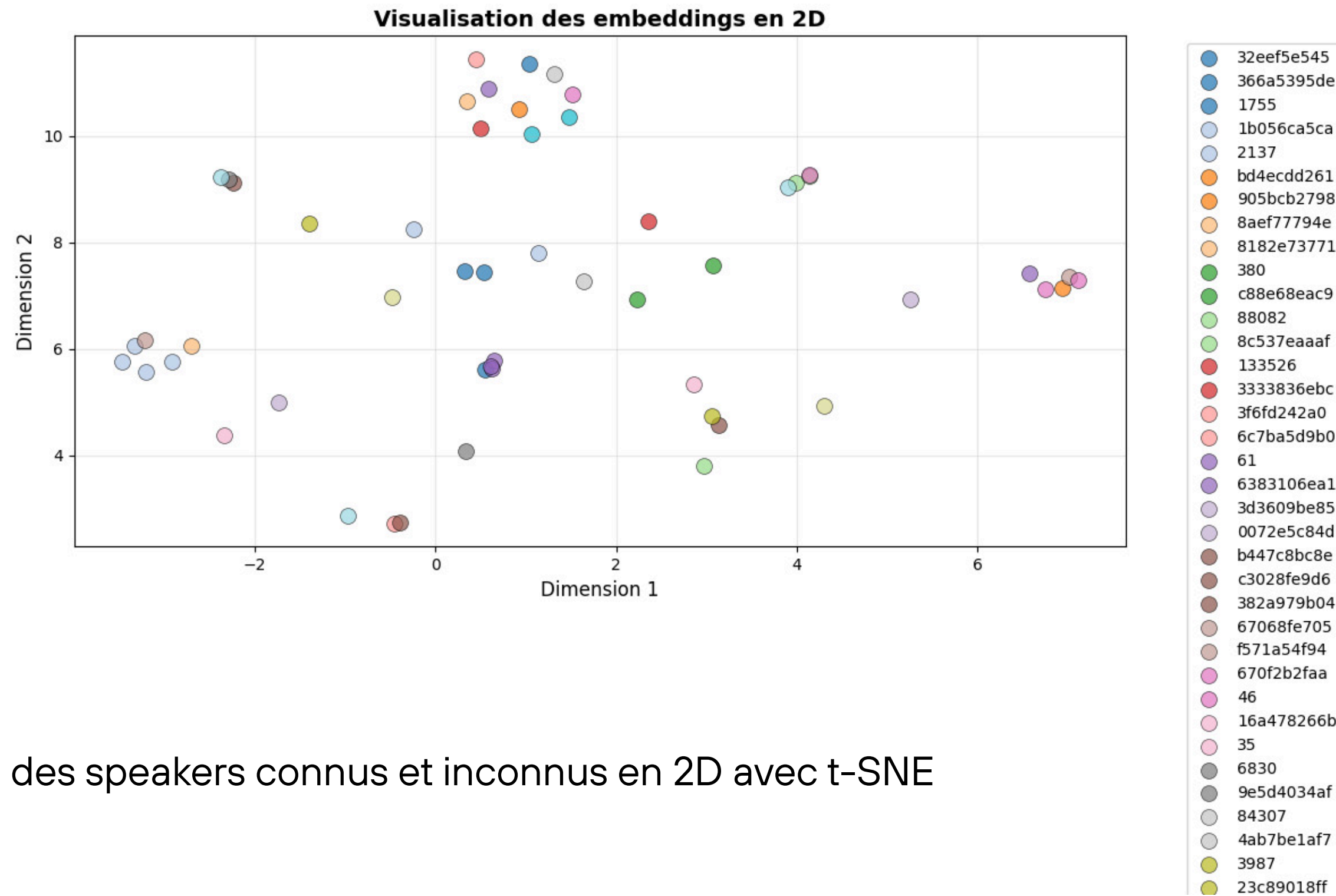
projection 2D des embeddings avec UMAP

m



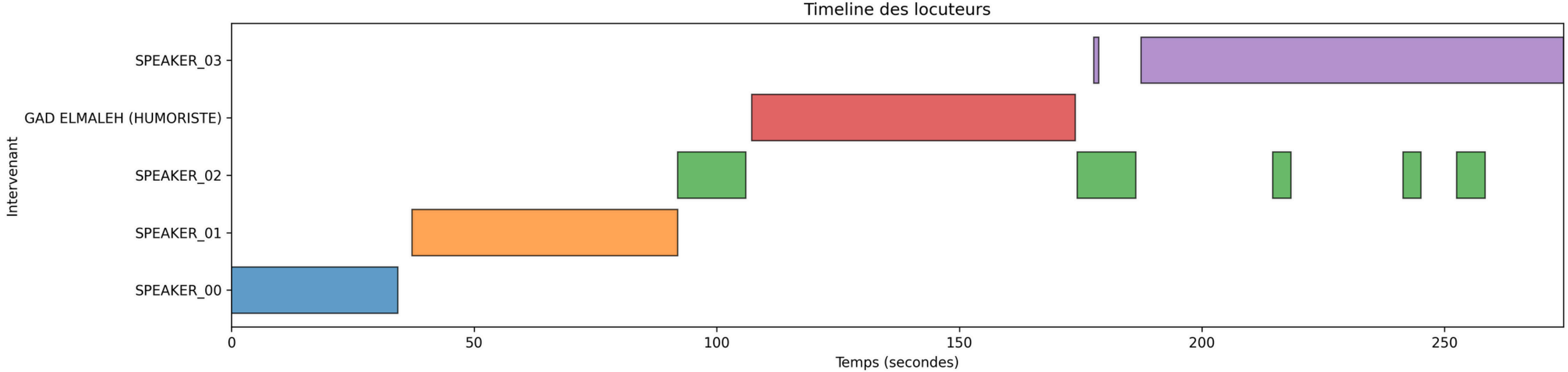
La même chose avec t-SNE

Résultats 2/2



Projection des speakers connus et inconnus en 2D avec t-SNE

Ouverture



Ouverture

- Utilisation des données de plus d'émissions
- Comparaison avec données de l'ARCOM
- Étude qualitative des résultats
- Identifier le nom des speakers inconnus grâce à un traitement NLP



Merci !

