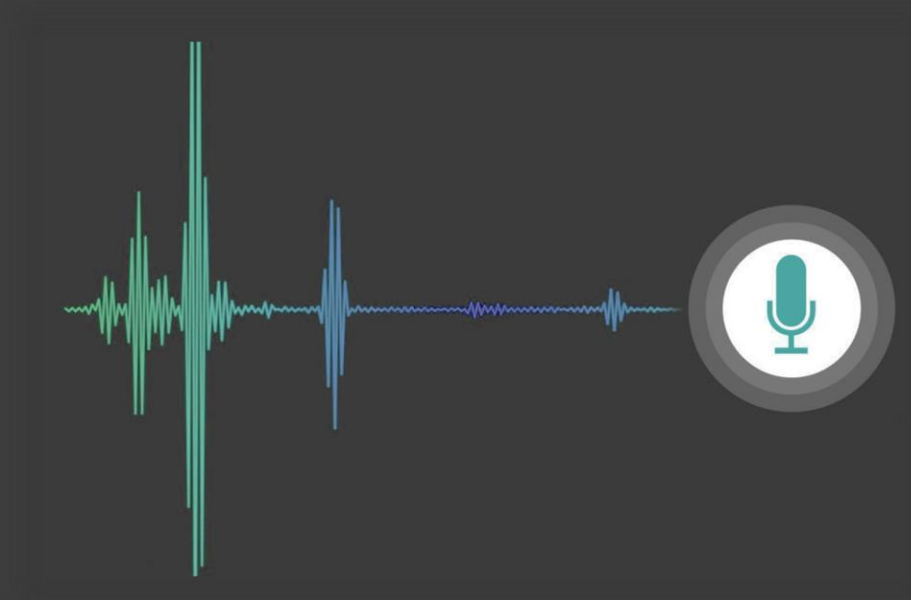


Projet 7

Développer une preuve de concept

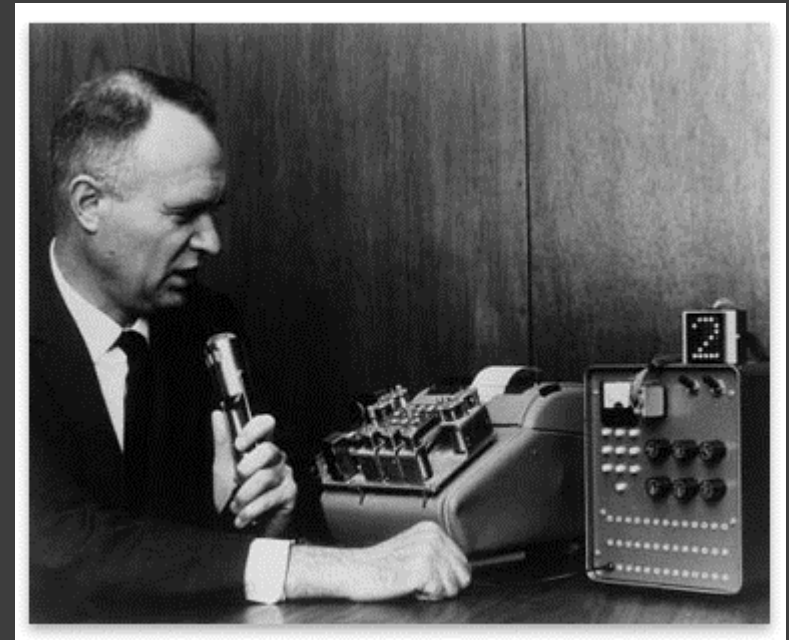
Thématique

La Reconnaissance vocale ou Automatic Speech Recognition (ASR)



Les Origines

- 1960 – Shoebox
- 1980 – Hidden Markov Models (HMMs)
- 1990 – usage grand publique (dragon)
- 2008 – Command vocal
- 2010 – Deep Neural Networks (DNN)

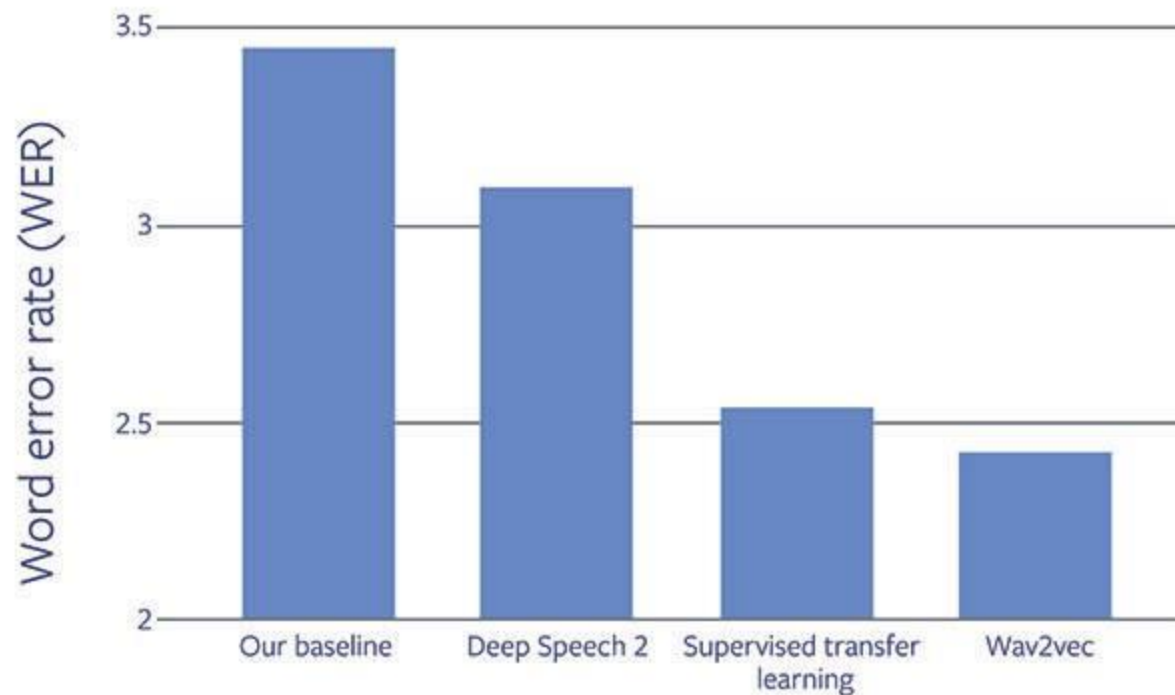


1961 — IBM Shoebox

Etat de l'art

Wav2vec

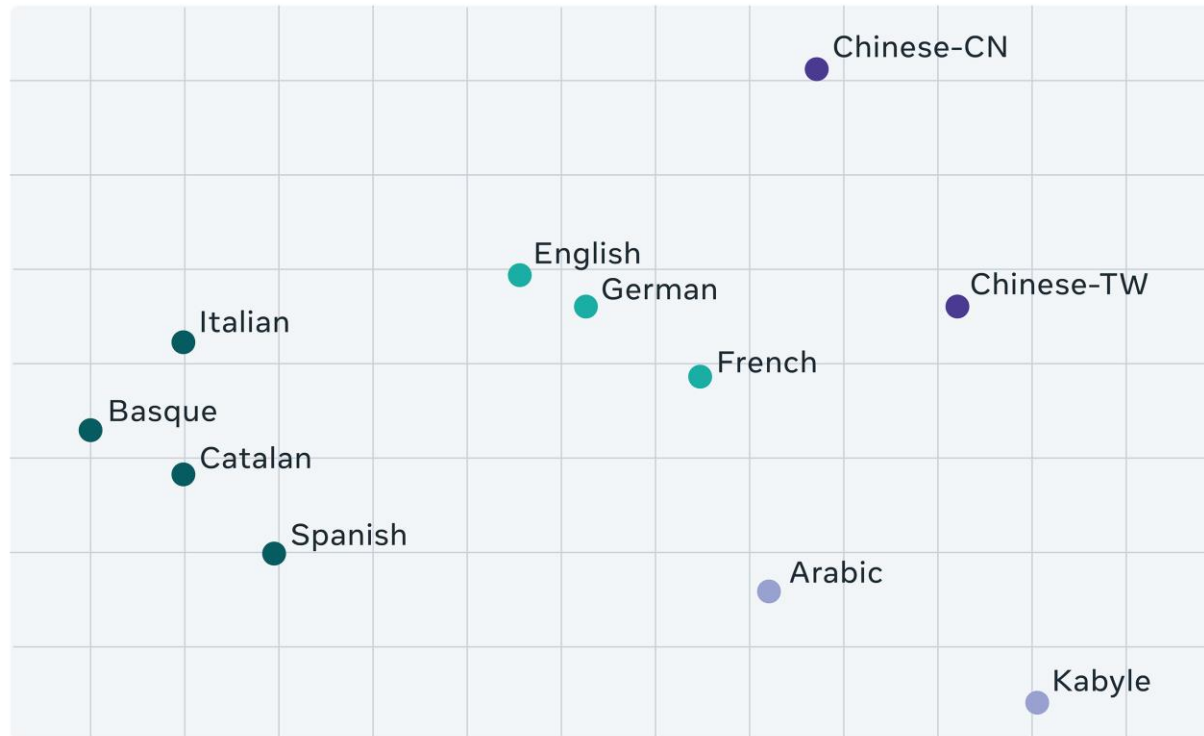
Les débuts de l'auto-apprentissage



Etat de l'art



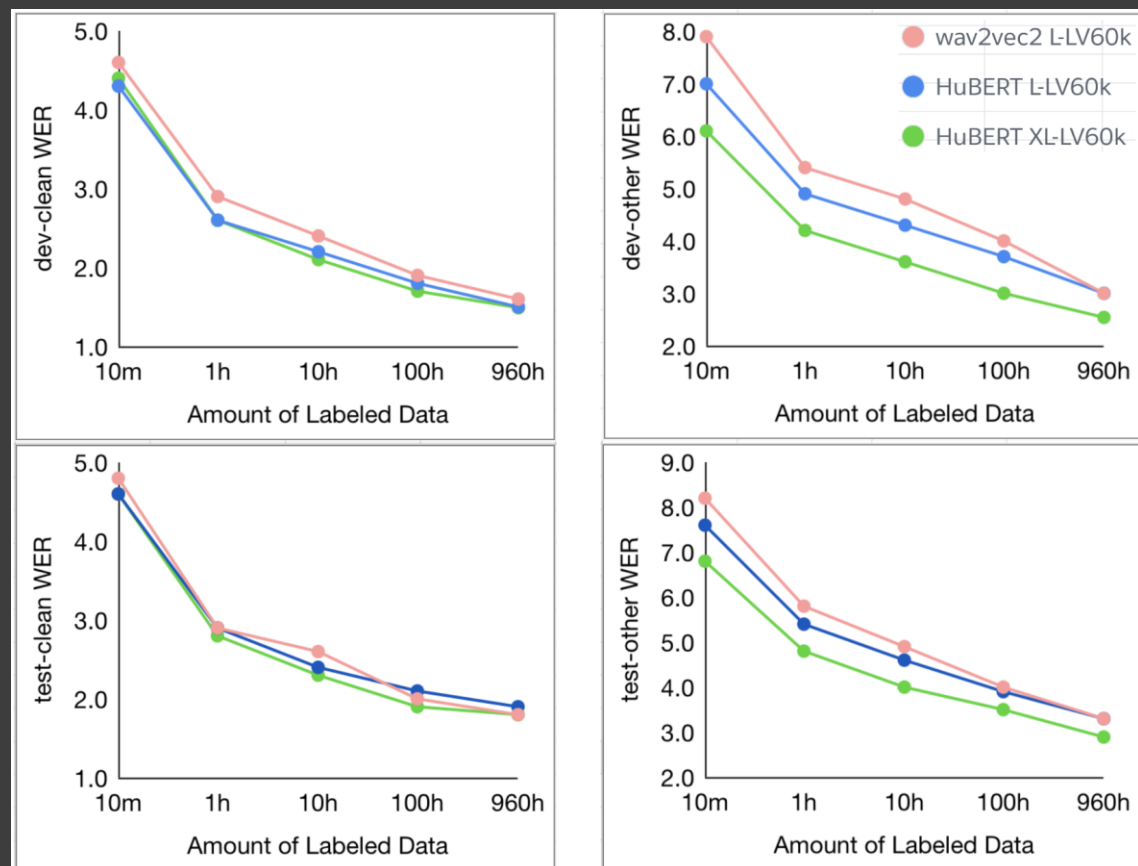
Meilleure reconnaissance des variables latentes



Etat de l'art

HuBERT

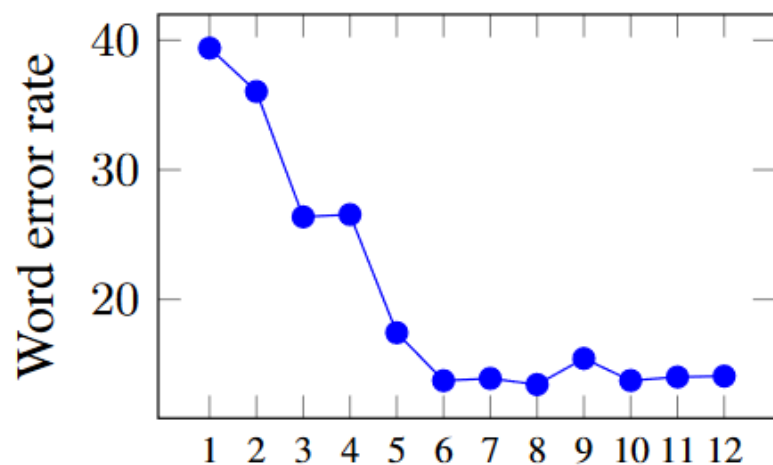
Apprentissage par donnée masquées



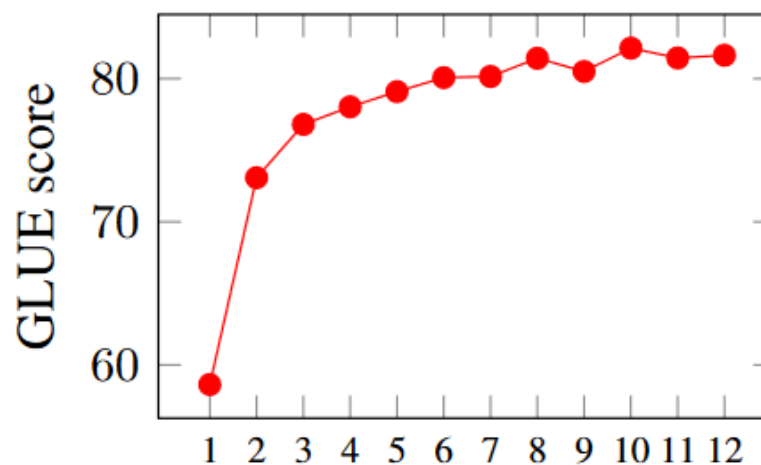
Etat de l'art



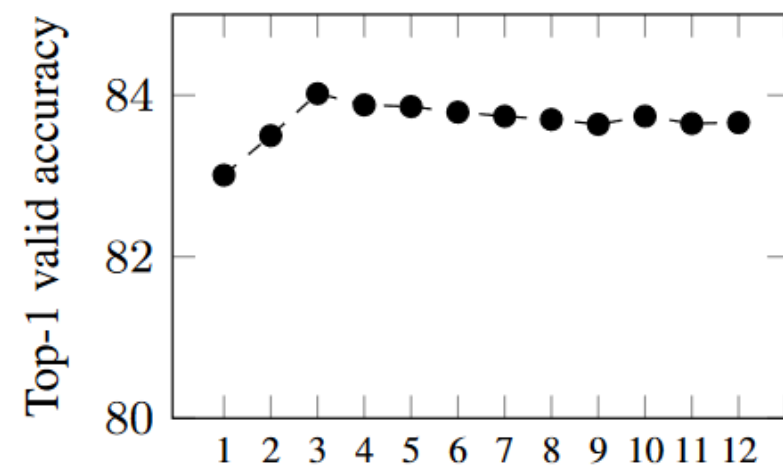
Apprentissage généralisé



(a) Speech



(b) NLP



(c) Vision

Jeu de donnée

Common
Voice

Heures validées

14 973



Heures enregistrées

20 217



Langues

93



Jeu de donnée



- 1 000 Heures de texte
- Extraits de livre audio
- Subdivision en niveau de difficulté

Métrique



WER

BUT THIS SHALL BANISH IT UTTERLY
BUT **THEY** **SOUBANISD** IT UTTERLY

WER : 1 manque + 2 erreurs + 0 ajout / 6 mots = 3/6 = 0.5

Métrique



BUT THIS SHALL BANISH IT UTTERLY
BUT THEY _____ SOUBANISD IT UTTERLY

CER : 5 manque + 3 erreurs + 3 ajout / 27 caractère = 11/27 = 0.4

Analyse des résultats

	WER	CER
Wav2vec clean	7.1%	2.7%
Data2vec clean	1.7%	0.4%
Wav2vec other	7.4%	3.3%
Data2vec other	3.0%	1.1%

Data2vec est en Moyenne plus de 3 fois
plus performant que Wav2vec