

TransAP - Transcripció automàtica de parla disàrtrica en català



Pol Pastells i Vilà

enginyer de recerca de lingüística computacional
**Centre de Llenguatge i Computació (CLiC),
Universitat de Barcelona**

1. SCRIBAL
2. AinaChallenge
3. Dades
4. Models
5. Plataforma
6. Demo
7. Conclusions

SCRIBAL: Transcriptor digital per a la docència

- Basat en el sistema WHISPER, d'OpenAI, que permet la **transcripció** en 99 llengües.
- També pot fer la **traducció** d'aquestes llengües a l'anglès



ADAPTACIÓ AL CATALÀ

Fine-tuning

Enregistraments propis
(professors UB)



ADAPTACIÓ TERMINOLÒGICA

Adaptació a dominis especialitzats
Utilització dels plans docents
Col·laboració amb Serveis Lingüístics UB



IMPLEMENTACIÓ

A LES AULES

Entorn web amigable
per a qualsevol
dispositiu

Transcriptor digital per a la docència

LLENGUA CATALANA

**TERMINOLOGIA
ESPECIALITZADA**

**DIFICULTATS AUDITIVES
I COGNITIVES**

Exportable a altres àmbits acadèmics/no acadèmics

**CONCURSOS,
OPOSICIONS**

**XERRADES,
CONFERÈNCIES**

CONGRESSOS

SCRIBAL



Projecte TransAP

SCRIBAL-Social



The poster features a blue background with abstract geometric shapes and arrows. At the top left is the BSC logo (Barcelona Supercomputing Center) and at the top right is the Generalitat de Catalunya logo. The main title 'Aina Challenge' is in large white font. Below it, the text describes a competition for AI demonstrators. At the bottom, it states the program's success with 22 awarded projects and a 1M€ investment.

BSC **Barcelona Supercomputing Center**
Centro Nacional de Supercomputación

Generalitat de Catalunya

Aina Challenge

Concurs de reptes per al desenvolupament de demostradors fent ús de recursos d'IA desenvolupats amb el Projecte Aina

Finalitza el programa amb 22 projectes premiats i una inversió d'1M€ que han aconseguit impulsar amb èxit la presència del català en la intel·ligència artificial.

Corpus



100 participants, 19.000 àudios curts, 25 hores de gravacions

Afectació	Nombre
Síndrome de Down	53
Paràlisi cerebral	34
Altres	13

Edat	Nombre
18-30	29
31-45	44
46-60	24
61-75	3

Dialecte	Nombre
Central (Barcelona)	66
Central (Girona)	16
Nord-Occidental	18

Gènere	Nombre
Dona	55
Home	43
Altres	2

Models

Recursos del projecte Aina

BSC-LT/

whisper-bsc-large-v3-cat (LT)

whisper-large-v3-ca-punctuated-3370h (LT-punct)

projecte-aina/

whisper-large-v3-ca-3catparla (Aina)

stt_ca-es_conformer_transducer_large (Conformer)

Infraestructura

Entrenament

- Marenostum 5*
- Servidor propi

Inferència

- Servidor Google Cloud
- Servidor propi

* A través del projecte EuroHPC JU EHPC-AIF-2025PG01-376

Adaptació inicial

Model	Mètode	Down	Paràlisi
LT	LoRA	32.50	44.10
	PiSSA	30.59	42.33
	FFT	29.40	40.46
LT-punct	LoRA	31.93	46.11
	PiSSA	30.28	42.32
	FFT	28.19	39.79
Aina	LoRA	33.70	46.70
	PiSSA	31.60	42.57
	FFT	29.13	40.64
Large-v3	LoRA	37.15	48.62
	PiSSA	34.55	45.72
	FFT	32.24	41.64
Large-v3-Turbo	LoRA	42.70	49.76
	PiSSA	36.12	46.41
	FFT	32.93	41.95

Metodologies

- Full Fine-tuning (FFT)
- Low-Rank Adaptation (LoRA)
- Principal Singular Values and Singular Vectors Adaptation (PiSSA)

Mètriques

- Mètrica WER (Word Error Rate)
- Avaluació qualitativa

Adaptació inicial

Model	Mètode	Down	Paràlisi
LT	LoRA	32.50	44.10
	PiSSA	30.59	42.33
	FFT	29.40	40.46
LT-punct	LoRA	31.93	46.11
	PiSSA	30.28	42.32
	FFT	28.19	39.79
Aina	LoRA	33.70	46.70
	PiSSA	31.60	42.57
	FFT	29.13	40.64
Large-v3	LoRA	37.15	48.62
	PiSSA	34.55	45.72
	FFT	32.24	41.64
Large-v3-Turbo	LoRA	42.70	49.76
	PiSSA	36.12	46.41
	FFT	32.93	41.95

Metodologies

- Full Fine-tuning (FFT)
- Low-Rank Adaptation (LoRA)
- Principal Singular Values and Singular Vectors Adaptation (PiSSA)

Millora entre
model d'OpenAI i
millor model d'Aina.

- Mean Perplexity (Error Rate)

- Avaluació qualitativa

Nous resultats

Model		WER		
	Síndrome de Down	Paràlisi Cerebral	Total	
Whisper, LT (FFT)	9.6	22.4	16.1	
Conformer (FFT)	16.0	27.6	21.8	

Re-anotació dades -> Reducció del ~50% del WER

Plataforma

SCRIBAL-SOCIAL: UN TRANSCRIPTOR D'ÀUDIO PER A PERSONES AMB PARLA NO FLUIDA

TECNOLOGIA PER A LA INCLUSIÓ I L'ACCESSIBILITAT.

SCRIBAL-SOCIAL ÉS UN TRANSCRIPTOR D'ÀUDIO CATALÀ PER A PERSONES AMB PARLA NO FLUIDA. EN CONCRET, ELS MODELS HAN ESTAT ENTRENATS AMB ENREGISTRAMENTS DE PERSONES AMB SÍNDROME DE DOWN O PARÀLISI CEREBRAL. SCRIBAL-SOCIAL HA ESTAT DESENVOLUPAT DINS DEL PROJECTE TRANSAP, FINANÇAT PER [AINACHALLENGE 2025](#).

↑ TRANSCRIPCIÓ
DE FITXERS

🎤 TRANSCRIPCIÓ
EN DIRECTE



ATURAR

AMAGAR SENYAL

Picture in Picture
(PiP)

EXEMPLE DE TRANSCRIPCIÓ EN DIRECTE

Plataforma

SCRIBAL-SOCIAL: UN TRANSCRIPTOR D'ÀUDIO PER A PERSONES AMB PARLA

TECNOLOGIA PER A LA

SCRIBAL-SOCIAL ÉS UN TRANSCRIPTOR D'ÀUDIO CATALÀ PER A
ESTAT ENTRENATS AMB ENREGISTRAMENTS DE PERSONES A
ESTAT DESENVOLUPAT DINS DEL PROJECTE

↑ TRANSCRIPCIÓ
DE FITXERS

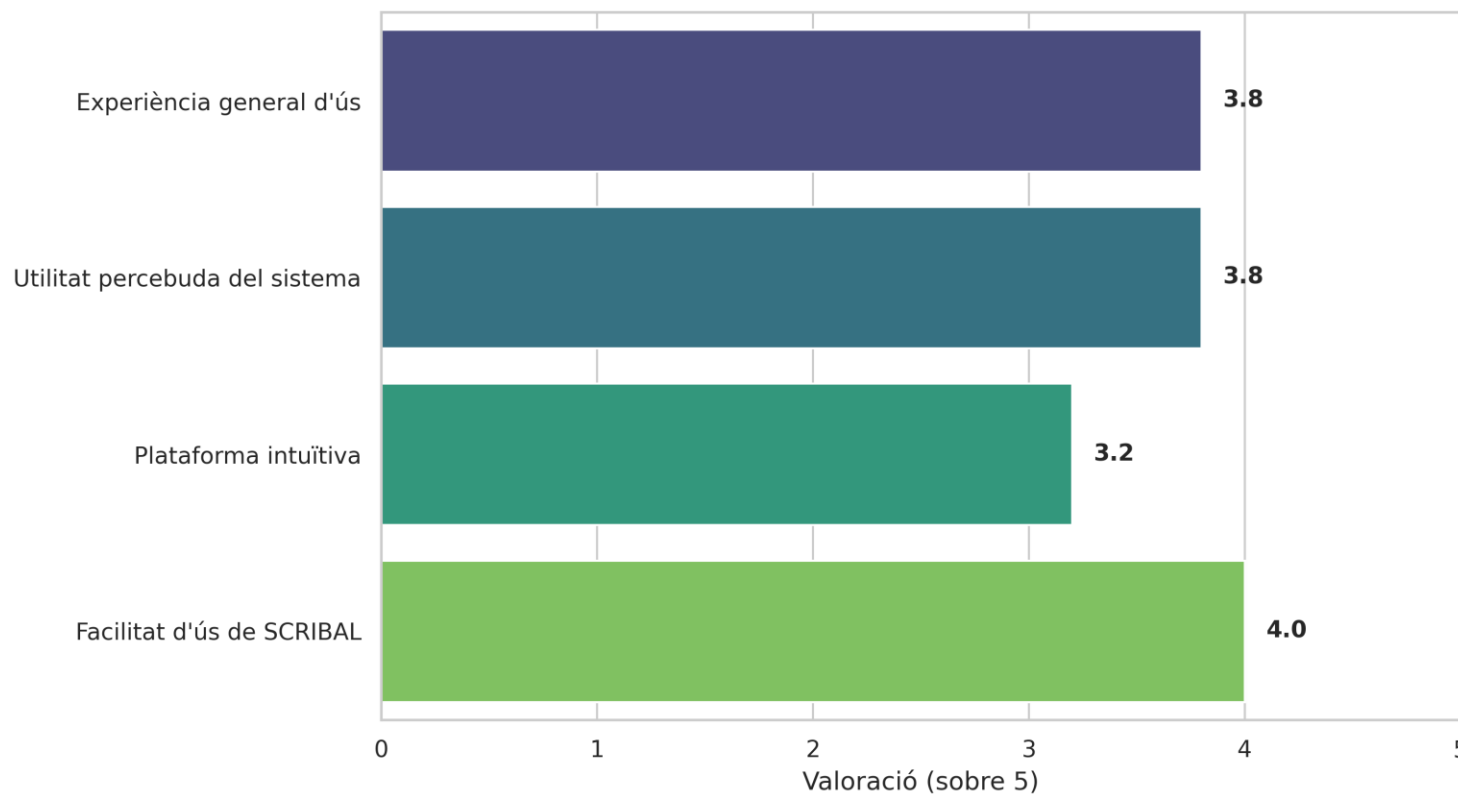


ATURAR

AMAGAR S

EXEMPLE DE TRANSCRIPCIÓ EN DIRECTE

Resultats de l'Experiència d'Usuari



DEMO

Conclusions

- 25 hores de parla disàrtrica en català
- Molt bons resultats per transcripció de frases intel·ligibles

Pròxims passos

Dades

Recopilació de contingut ja creat a internet.

Models

Personalització per a individus concrets (en curs)

TransAP - Transcripció automàtica de parla disàrtrica en català

Pol Pastells