CIDAI Centre of Innovation for Data tech and Artificial Intelligence

Socis promotors

Generalitat de Catalunya

Ajuntament de Barcelona

BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación

CVC Centre de Visió per Computador

eurecat Centre Tecnològic de Catalunya

i2cat

HUAWEI

Microsoft

NTT DaTa

SAP

sdg group

UPC Universitat Politècnica de Catalunya BarcelonaTech Intelligent Data Science and Artificial Intelligence Research Center IDEAI

Membres

AMB

ATM Àrea de Barcelona Autoritat del Transport Metropolità

Caixa Enginyers Banca Cooperativa

CETAQUA Centro Tecnológico del Agua

3cat

CRUÏLLA

PRESENTACIÓ PROVES DE CONCEPTE DE DADES I IA

16 maig 2024

www.cidai.eu

CIDAI Centre of Innovation
for Data tech
and Artificial Intelligence

MATXA, the first speech open-source solution to support the different Catalan varieties - Barcelona Supercomputing Center (BSC)

MATXA, the first speech open-source solution to support the different Catalan varieties –BSC

Martí Llopart Font

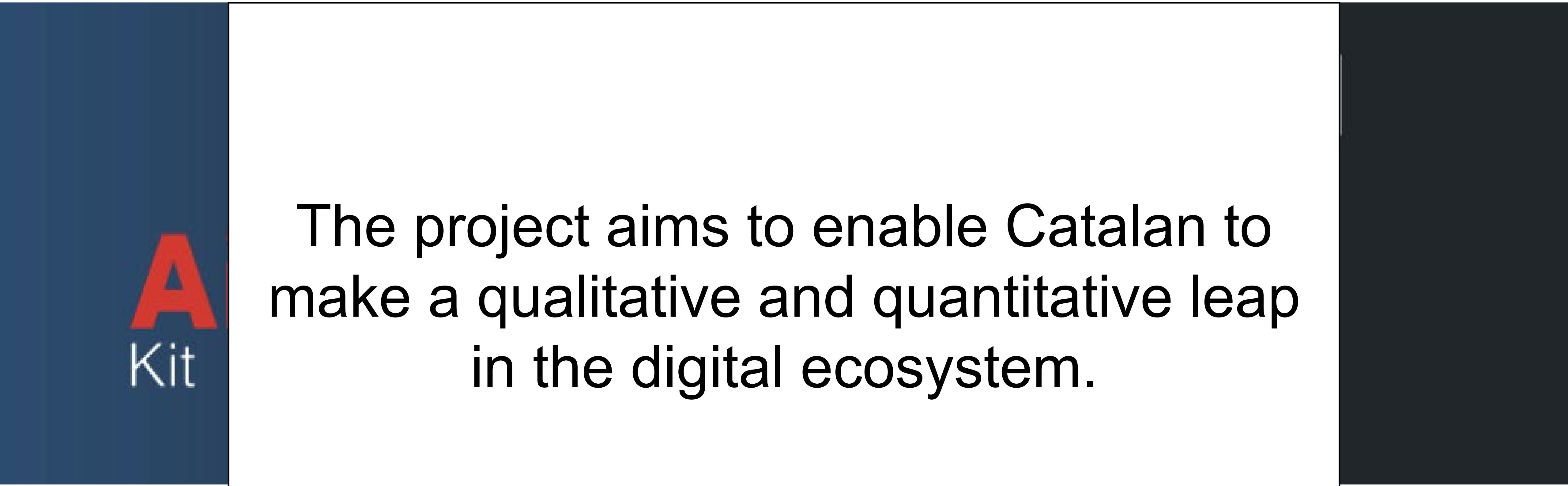
Direcció Innovació Digital

Barcelona Supercomputing Center (BSC)

MATXA

The first speech open-source solution to suport the different Catalan varieties - BSC

Who are we?



The project aims to enable Catalan to make a qualitative and quantitative leap in the digital ecosystem.

MATXA

The first speech open-source solution to suport the different Catalan varieties - BSC

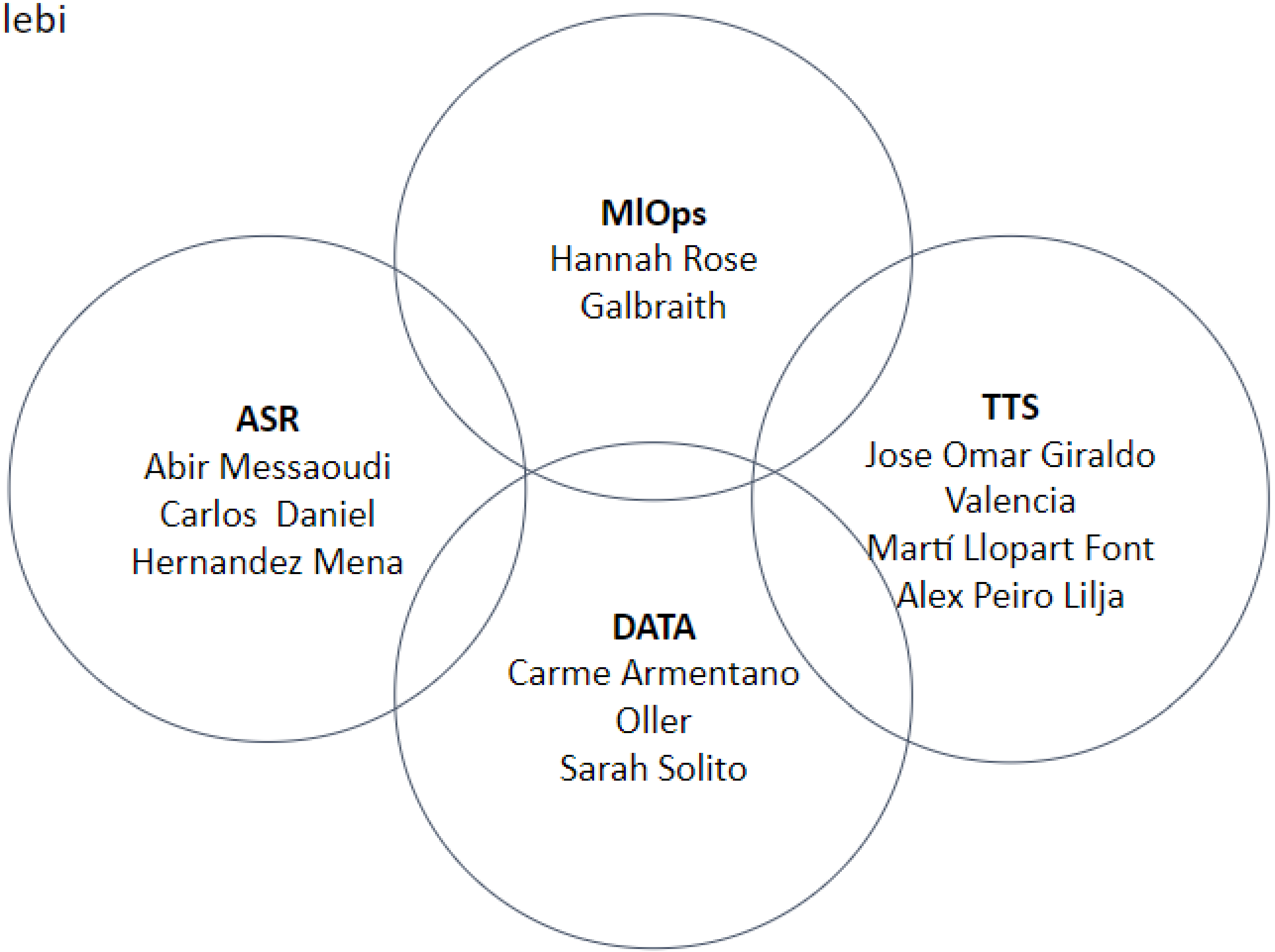
Who are we?

Coordination: Baybars Kulebi

Speech Research

Coordination:
Javier Hernando

Federico Costa
Marc Casals Salvador



MlOps
Hannah Rose
Galbraith

ASR
Abir Messaoudi
Carlos Daniel
Hernandez Mena

TTS
Jose Omar Giraldo
Valencia
Martí Llopart Font
Alex Peiro Lilja

DATA
Carme Armentano
Oller
Sarah Solito

MATXA

The first speech open-source solution to suport the different Catalan varieties - BSC

Who am I ?

Martí Llopart, BSC – Language Technologies Unit – TTS from Speech

BEng Biomedical Engineering – First Class Honours

Queen Mary University of London

MEM Engineering Project Management

UPC

Published in the renowned journal of Biophysical Reviews, cited by Nature

Published by Databricks

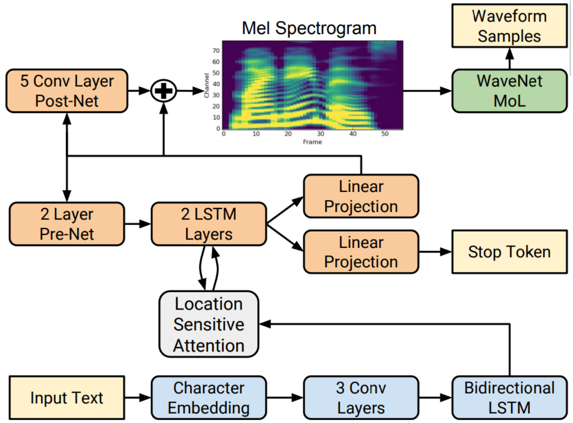
MATXA

The first speech open-source solution to suport the different Catalan varieties - BSC

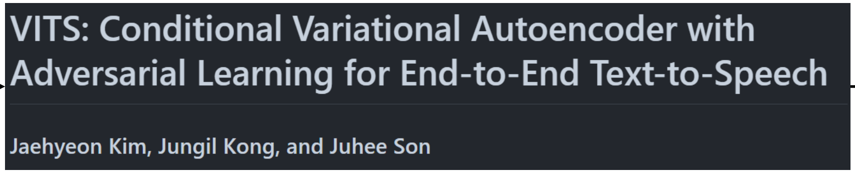
Where do we come from?

CATOTRON

SÍNTESI DE LA PARLA OBERTA I LLIURE EN CATALÀ

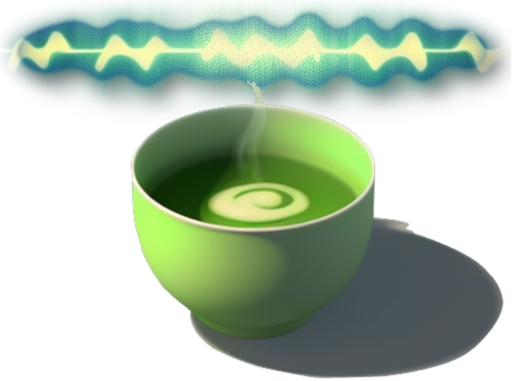


TTS


VITS: Conditional Variational Autoencoder with Adversarial Learning for End-to-End Text-to-Speech

Jaehyeon Kim, Jungil Kong, and Juhee Son

Matcha-TTS



MATXA

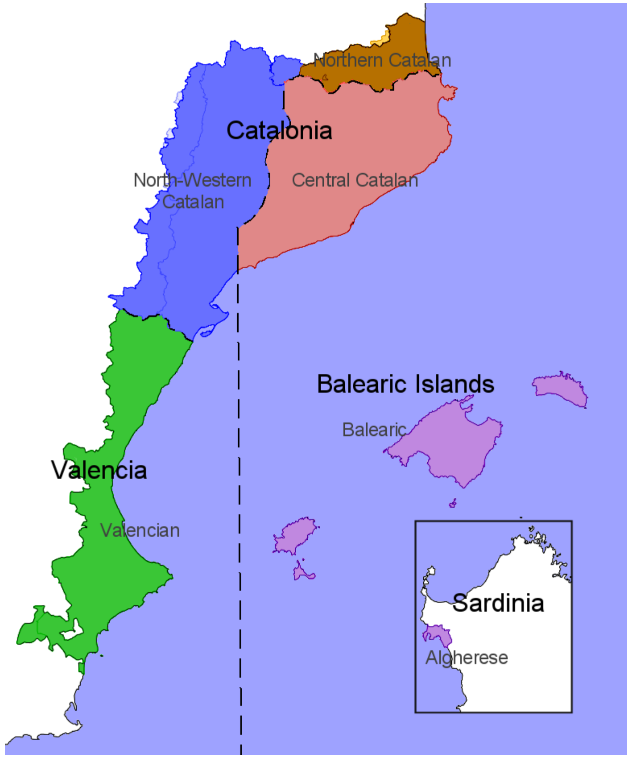
The first speech open-source solution to suport the different Catalan varieties - BSC

Summary

What’s Matxa?

It is the first multispeaker, multidialectal neural TTS model, and comes together with the vocoder model alVoCat to generate high quality and expressive speech efficiently in four Catalan dialects:

- Balear
- Central
- North-Occidental
- Valencian

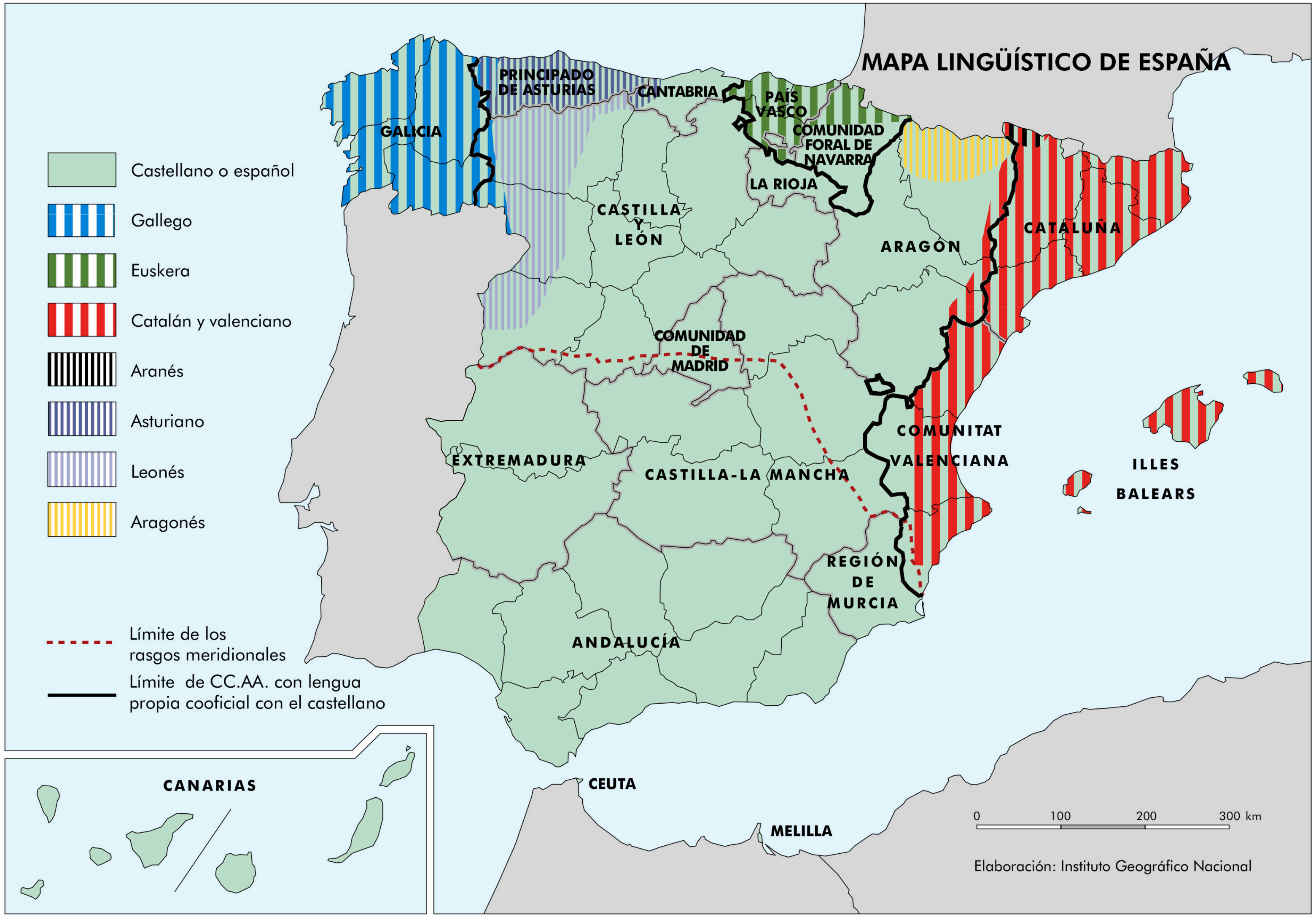


MATXA

The first speech open-source solution to suport the different Catalan varieties - BSC

TTS in minority languages

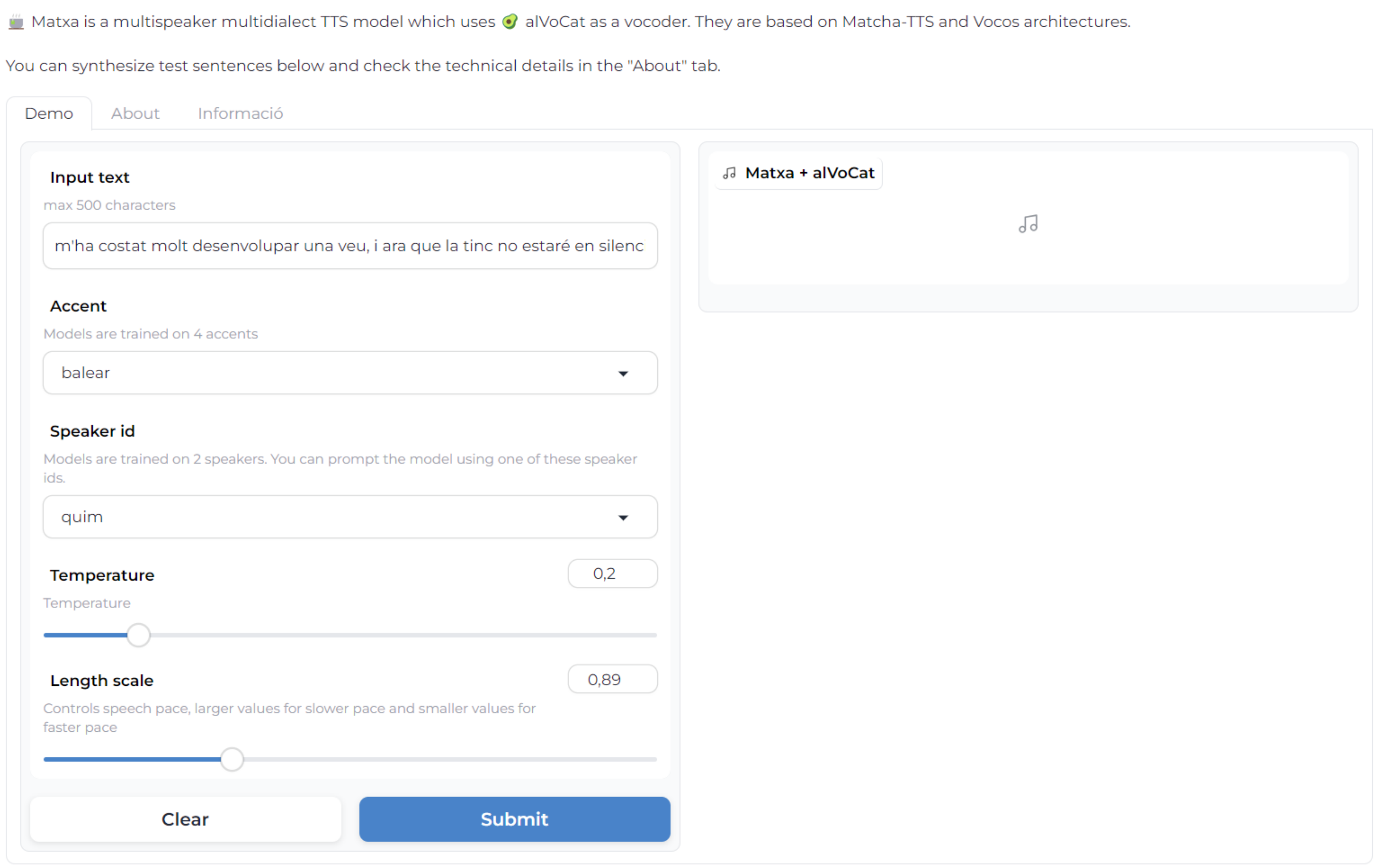
More efforts are needed to democratise these solutions.



MATXA

The first speech open-source solution to suport the different Catalan varieties - BSC

What we intend to do



Matxa is a multispeaker multidialect TTS model which uses alVoCat as a vocoder. They are based on Matcha-TTS and Vocos architectures.

You can synthesize test sentences below and check the technical details in the "About" tab.

Input text

max 500 characters

m'ha costat molt desenvolupar una veu, i ara que la tinc no estaré en silenc

Accent

Models are trained on 4 accents

balear

Speaker id

Models are trained on 2 speakers. You can prompt the model using one of these speaker ids.

quim

Temperature 0,2

Temperature

Length scale 0,89

Controls speech pace, larger values for slower pace and smaller values for faster pace

Objectives

- Developing Natural-Sounding TTS Synthesis for Catalan Dialects
- Seamless Integration with the administration for Visually Impaired Assistance
- User-Friendly Interface and Model Download

CIDAI Centre of Innovation for Data tech and Artificia

Matxa

varieties - BSC

Current

COST

Quantity

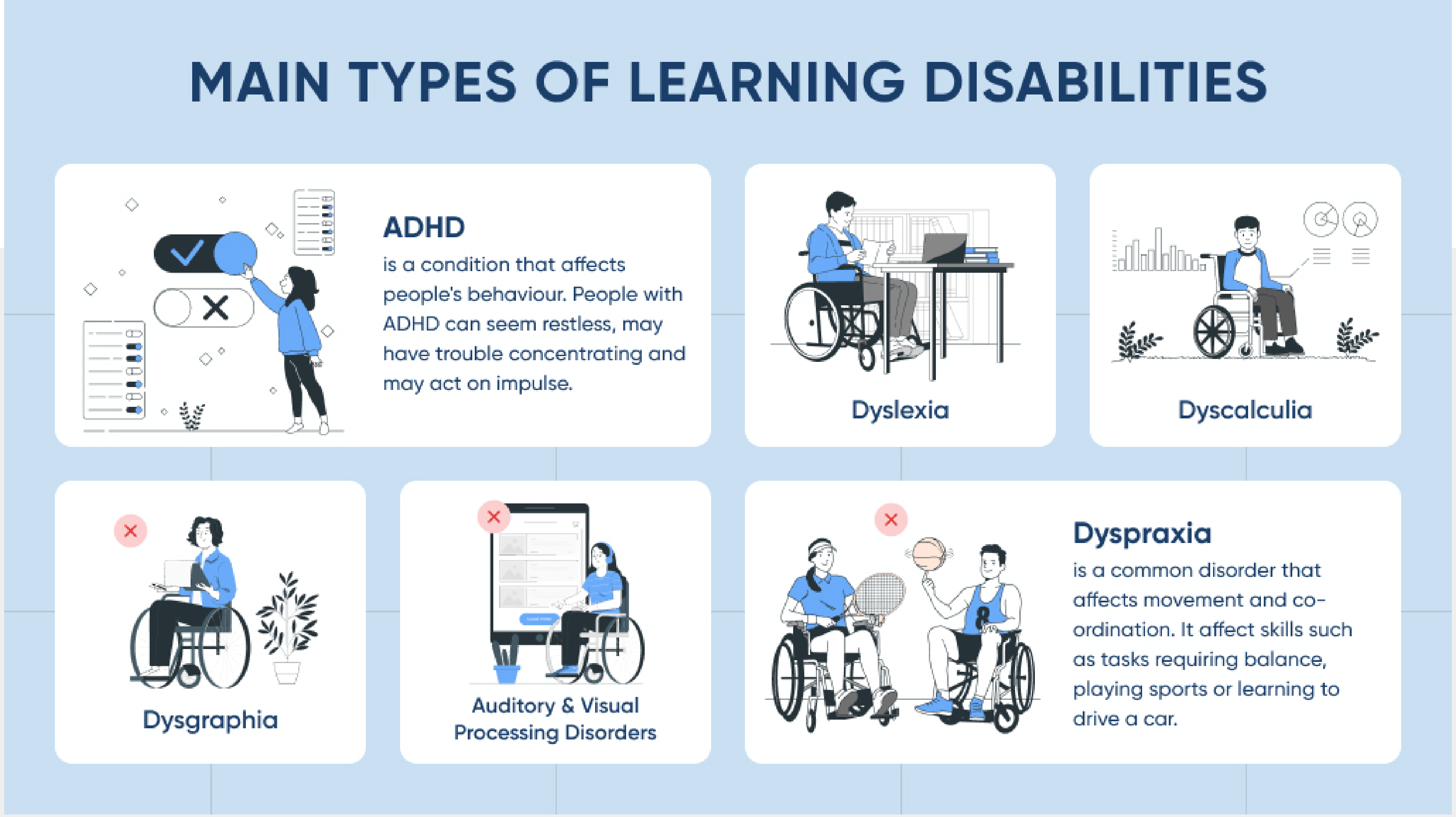
Standardized

Centre of Innovation
for Data tech
and Artificial Intelligence

Matxa

The first speech open-source solution to suport the different Catalan varieties - BSC

What’s the problem to be solved?



MAIN TYPES OF LEARNING DISABILITIES

ADHD
is a condition that affects people's behaviour. People with ADHD can seem restless, may have trouble concentrating and may act on impulse.

Dyslexia

Dyscalculia

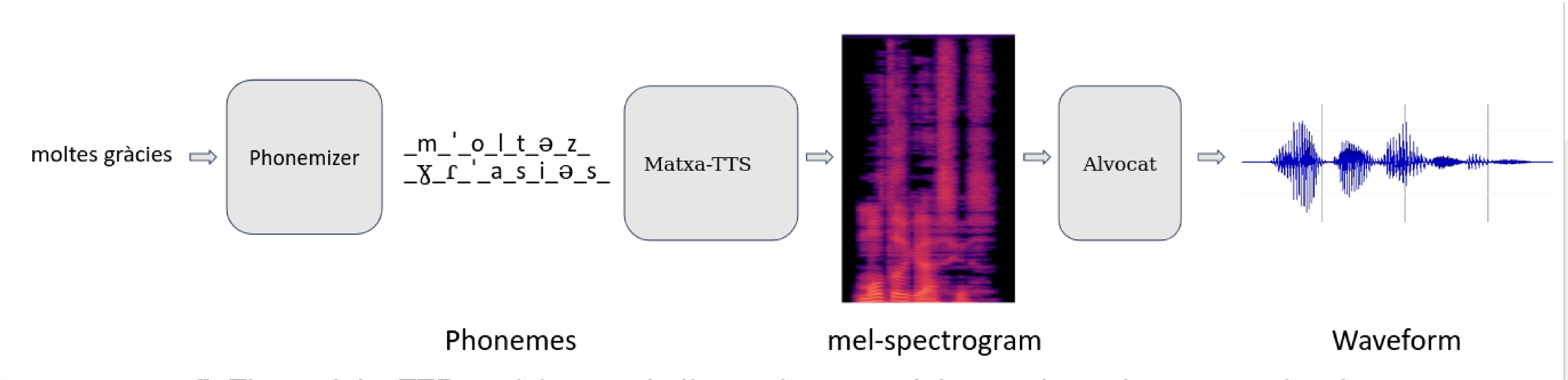
Dysgraphia

Auditory & Visual Processing Disorders

Dyspraxia
is a common disorder that affects movement and co-ordination. It affect skills such as tasks requiring balance, playing sports or learning to drive a car.

Matxa

The first speech open-source solution to suport the different Catalan varieties - BSC



2- The matcha-TTS model converts these phonemes into a mel spectrogram, a visual representation of the spectrum of frequencies of a sound over time.

3- This spectrogram is then fed into our adaptation of the Vocos vocoder, which synthesizes the speech waveform.

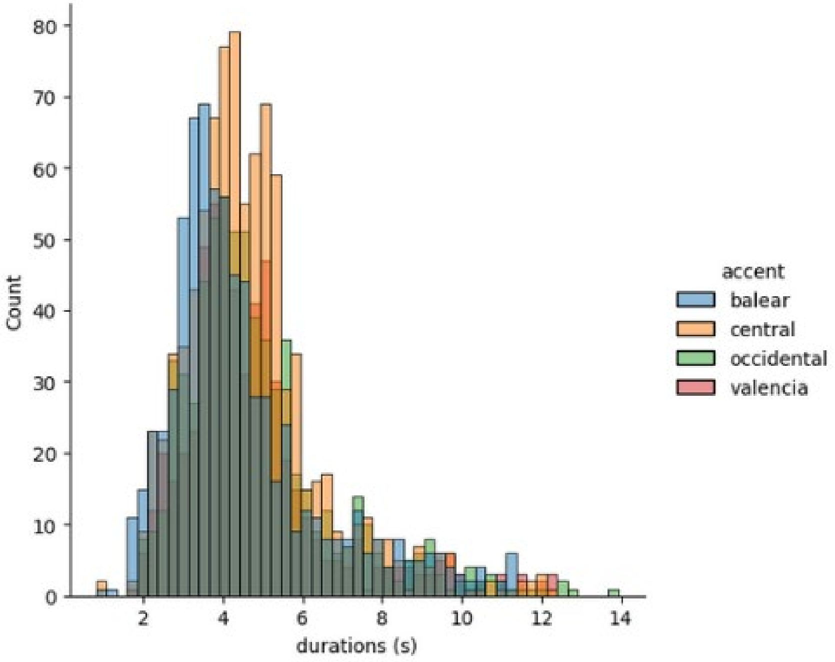
Matxa

The first speech open-source solution to suport the different Catalan varieties - BSC

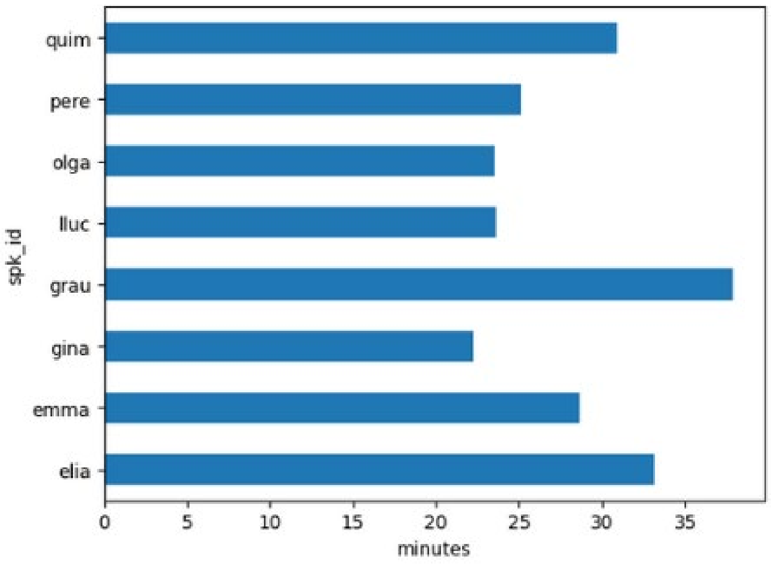
• Multi-accent data:

- 3.5h of studio recordings
- Two voices per accents (female/male)
- On average, 25 min per speaker.

Distribution of utterance durations per accent



Total durations per speaker



• Multi-accent data:

In the following example extracted from our eSpeak, clear phonetic differences can be observed:

Original sentence: “Volem sentir la teva veu perquè és molt important”

Balearic: volˈɛn səntˈi lə tˈevə vˈɛw pəɾkˈə ˈəz mˈolt impoɾtˈant

Central: bulˈɛm səntˈi lə tˈeβə βˈɛw pərkˈɛ ˈez mˈol impurtˈan

North-Western: bolˈem sentˈi la tˈewɛ βˈew perkˈe ˈez mˈol importˈan

Valencian: volˈem sentˈiɾ la tˈewa vˈew peɾkˈe ˈez mˈolt impoɾtˈant

Matxa

The first speech open-source solution to suport the different Catalan varieties - BSC

Balearic: 🔊

Central: 🔊

North-Western: 🔊

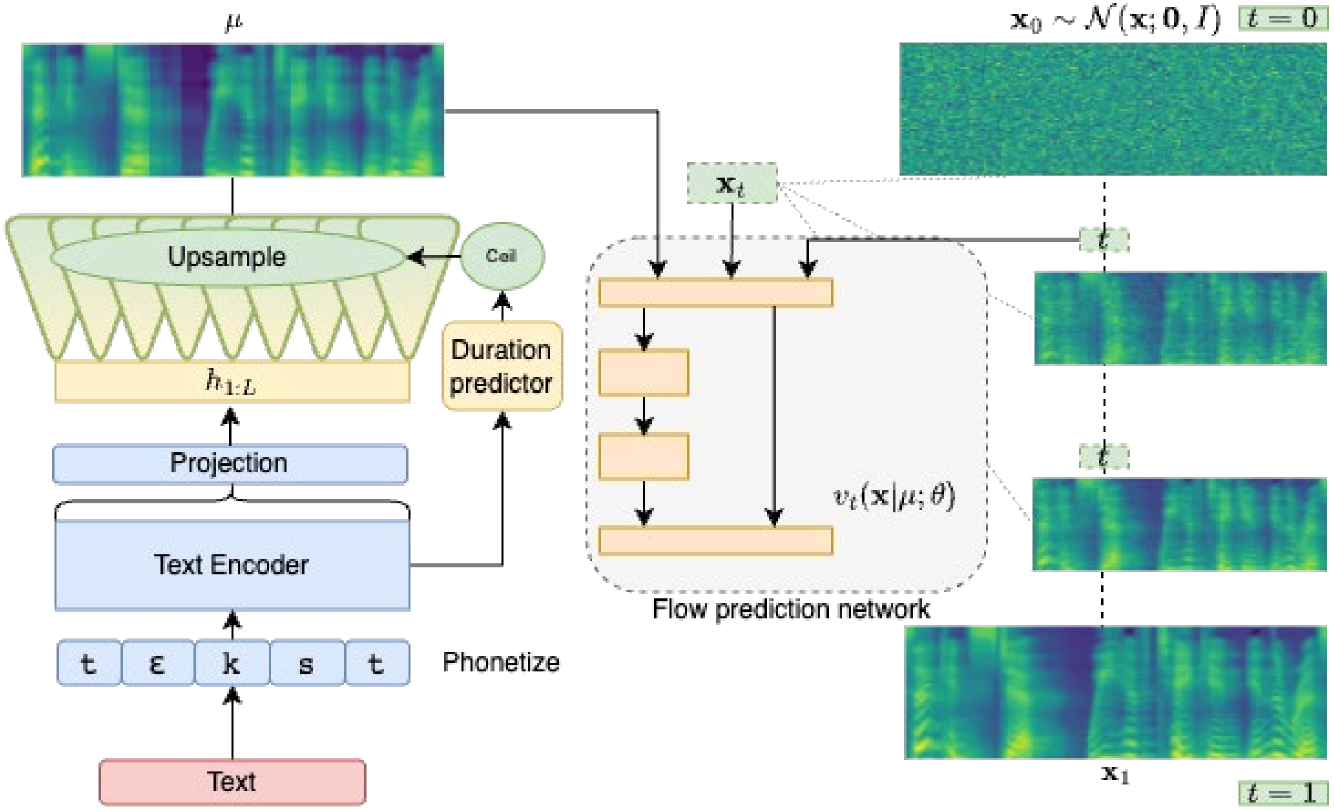
Valencian: 🔊

Matxa

The first speech open-source solution to suport the different Catalan varieties - BSC

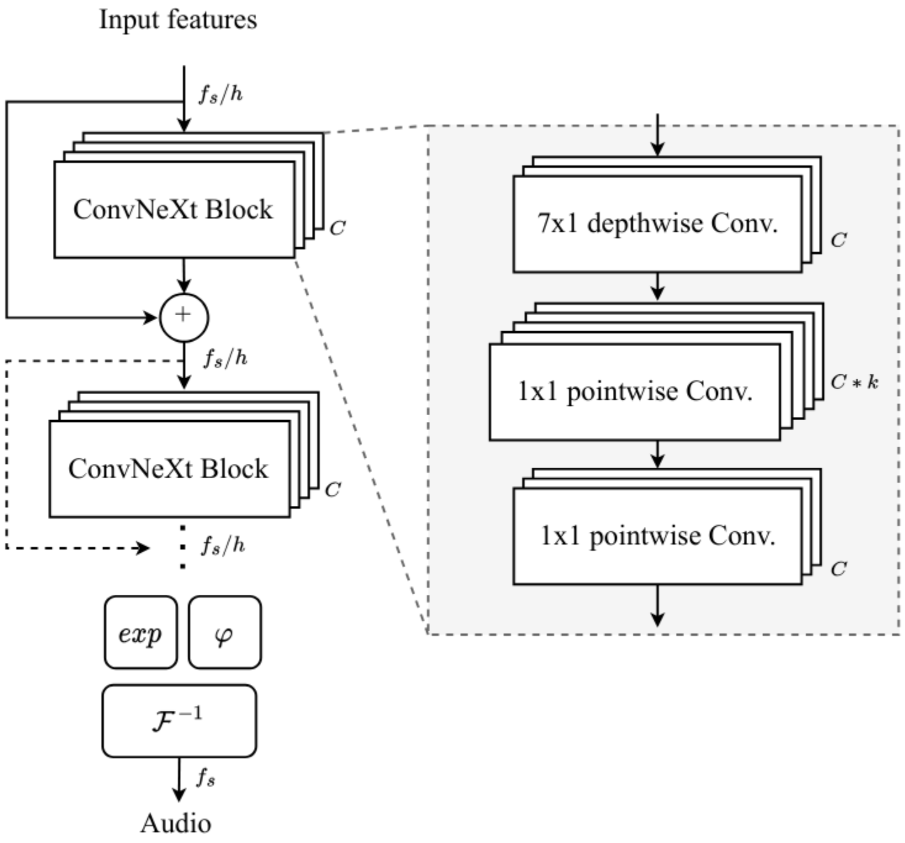
• Model architectures:

Matxa is based on Matcha-TTS[1], a non-autoregressive encoder-decoder model designed for fast acoustic modelling.



• Model architectures:

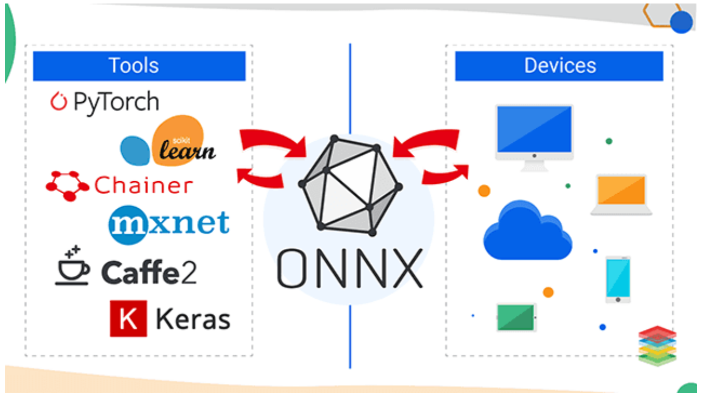
AlVoCat is an adapted version of the recently published vocoder named Vocos[3]. It is a fast neural vocoder designed to synthesise audio waveforms from acoustic features.

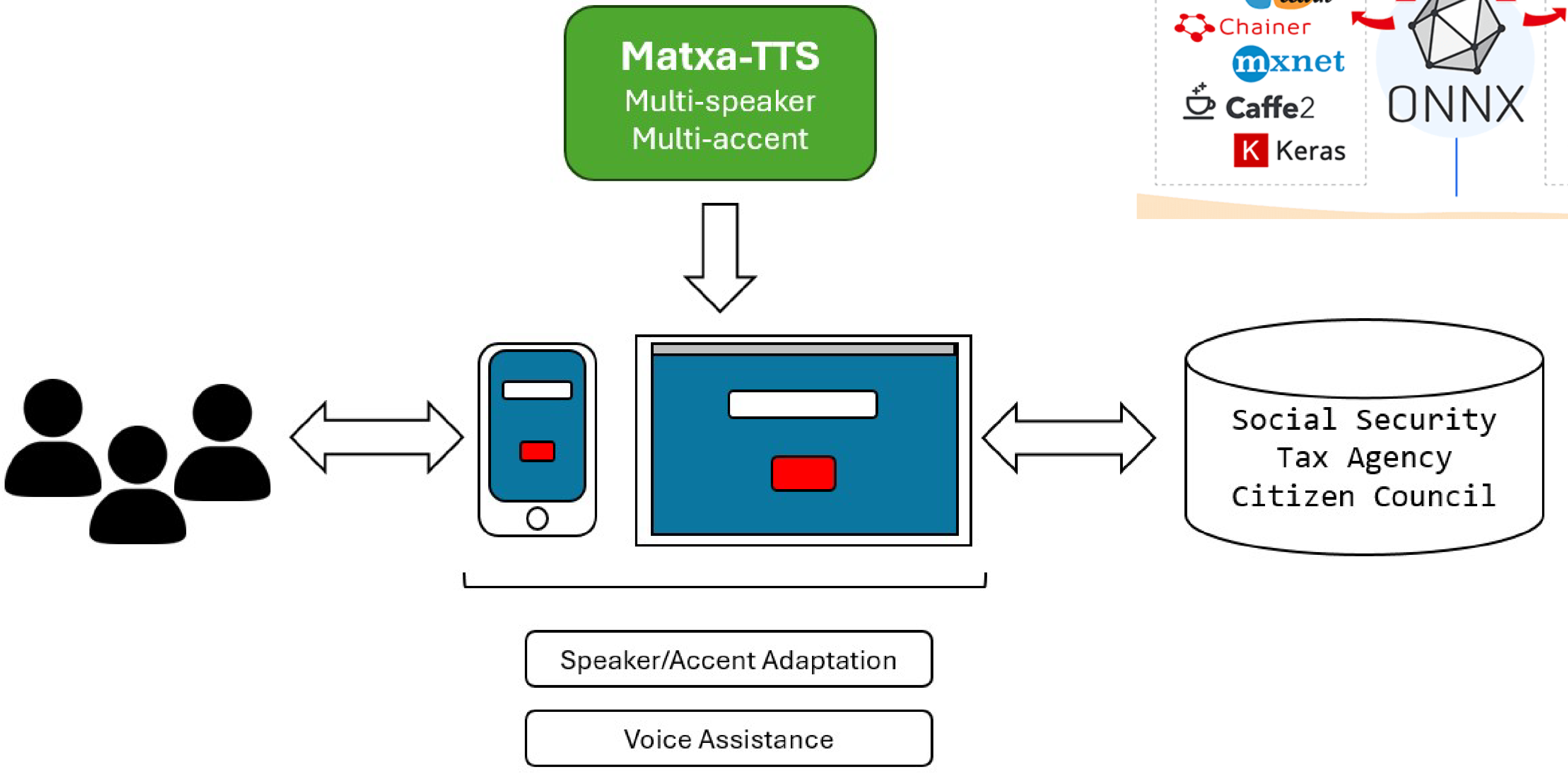


Matxa

The first speech open-source solution to suport the different Catalan varieties - BSC

- Implementation





Matxa-TTS
Multi-speaker
Multi-accent

Social Security
Tax Agency
Citizen Council

Speaker/Accent Adaptation

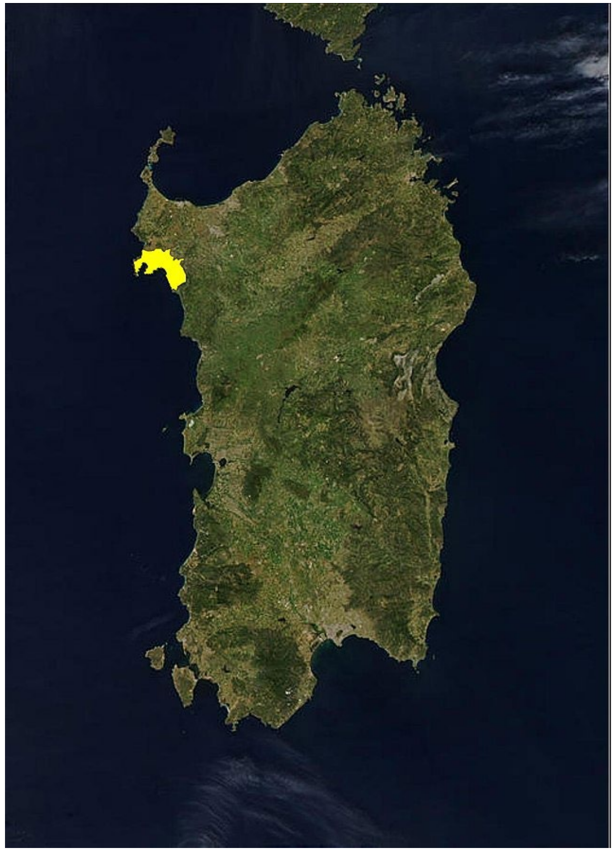
Voice Assistance

Matxa

The first speech open-source solution to suport the different Catalan varieties - BSC

Impact of the solution

Septentrional +



Centre of Innovation for Data tech and Artificial Intelligence

Matxa

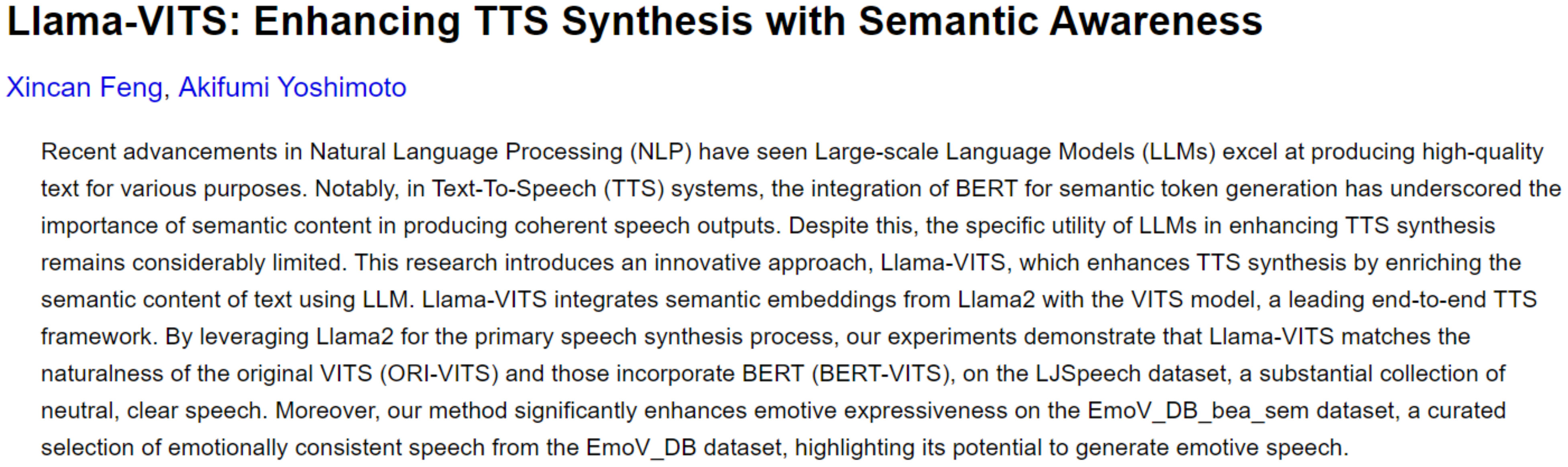
The first speech open-source solution to suport the different Catalan varieties - BSC

Impact of the solution

TEXT_Models

Encoders / Decoders models, foundational, pretrained or fine-tuned

projecte-aina/FLOR-6.3B



Llama-VITS: Enhancing TTS Synthesis with Semantic Awareness

Xincan Feng, Akifumi Yoshimoto

Recent advancements in Natural Language Processing (NLP) have seen Large-scale Language Models (LLMs) excel at producing high-quality text for various purposes. Notably, in Text-To-Speech (TTS) systems, the integration of BERT for semantic token generation has underscored the importance of semantic content in producing coherent speech outputs. Despite this, the specific utility of LLMs in enhancing TTS synthesis remains considerably limited. This research introduces an innovative approach, Llama-VITS, which enhances TTS synthesis by enriching the semantic content of text using LLM. Llama-VITS integrates semantic embeddings from Llama2 with the VITS model, a leading end-to-end TTS framework. By leveraging Llama2 for the primary speech synthesis process, our experiments demonstrate that Llama-VITS matches the naturalness of the original VITS (ORI-VITS) and those incorporate BERT (BERT-VITS), on the LJSpeech dataset, a substantial collection of neutral, clear speech. Moreover, our method significantly enhances emotive expressiveness on the EmoV_DB_bea_sem dataset, a curated selection of emotionally consistent speech from the EmoV_DB dataset, highlighting its potential to generate emotive speech.

Acknowledgments

The development of LaFresCat dataset, and the neural network models Matxa and alvoCat has been possible thanks to the financing by the Government of Catalonia through the Aina project.

Aina

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

El projecte Aina ha estat impulsat i finançat per la

Generalitat de Catalunya

Nota Legal

Copyright © 2024 Projecte Aina TECH

Thank you for your attention!

Any requests?

CIDAI

Centre of Innovation
for Data tech
and Artificial Intelligence

Socis promotors

Generalitat de Catalunya

Ajuntament de Barcelona

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

CVC Centre de Visió per Computador

eurecat
Centre Tecnològic de Catalunya

i2cat

HUAWEI

Microsoft

NTT DATA

SAP

sdg group

UPC Universitat Politècnica de Catalunya BarcelonaTech
Intelligent Data Science and Artificial Intelligence Research Center
IDEAI

Membres

AMB

ATM Àrea de Barcelona
Autoritat del Transport Metropolità

Caixa Enginyers
Banca Cooperativa

CETAQUA
Centro Tecnológico del Agua

3 Catràdio 3cat

CRUÏLLA

www.cidai.eu