

Tecnologías del lenguaje en la frontera de la IA
Nuevo centro de referencia en tecnologías del lenguaje:

HiTZ

Hizkuntza **T**eknologiako **Z**entroa

Basque Centre for Language Technology

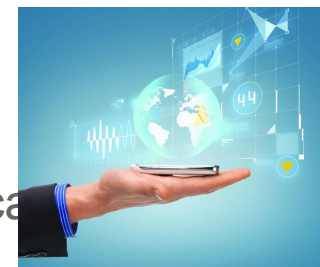


HiTZ

Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology



Sociedad de la información y tecnologías del lenguaje



- PLN en el top 10 de las tendencias tecnológicas estratégicas
- ~90 % de información digital no estructurada
 - Texto (multilingüe), audio, video, ...
- Transformación digital, amenaza para las lenguas minoritarias



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

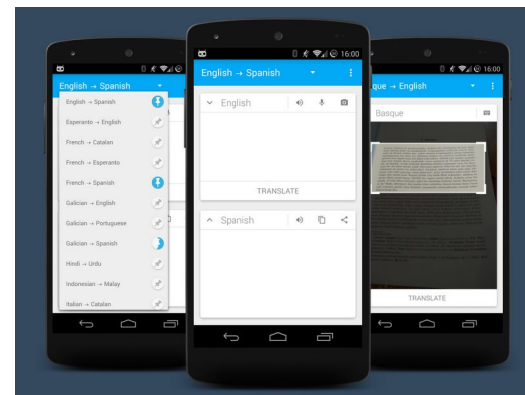
HiTZ

Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology



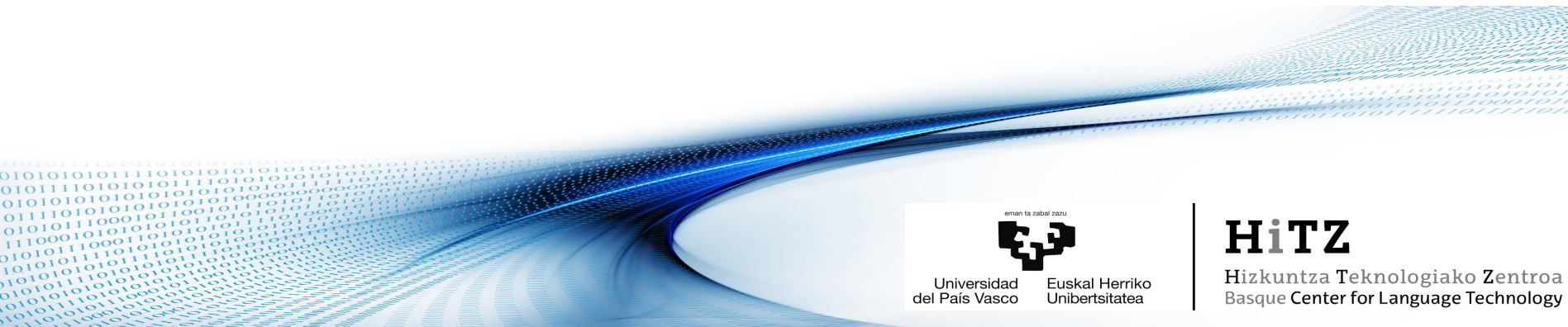
¿Qué es el PLN?, ¿Cuándo lo usamos?

- *Comprender y generar* texto, voz en un mundo global y multilingüe
 - Extracción de Información
 - Recuperación de Información
 - Chatbots
 - Traducción Automática
- Ejemplos
 - Efectos adversos de medicamentos
 - Medios y medios sociales
 - Texto ⇔ Imágen
 - Siri, Alexa, ...



Cada vez que hacemos una búsqueda en el ordenador, cada vez que el corrector nos corrige, cada vez que usamos la traducción automática...

Está es nuestra vida cotidiana!



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

HiTZ

Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

Nuevo centro de referencia en tecnologías del lenguaje:

HiTZ

Hizkuntza Teknologiako Zentroa
Basque Centre for Language Technology

Se crea para activar **la transferencia efectiva de tecnología lingüística** a nuestra sociedad, empresas, e instituciones

La tecnología lingüística es **un sector clave** para el desarrollo económico y social (Big Data e Inteligencia Artificial)

Partiendo de la experiencia de los **Grupos Ixa y Aholab**, servirá para **incorporar a otros agentes** del sector

Las opciones abiertas por el **Plan de Impulso de las Tecn. del Lenguaje** han hecho más perentorio el impulso de este centro.





HiTZ: Nuevo centro en tecnologías del lenguaje

- Promovido por la UPV/EHU (aprobado el 27 de septiembre de 2018)
 - Siguiendo modelos de centros en vigor: PIE, CFAA
- **Objetivo: potenciar formación, investigación, transferencia e innovación en las tecnologías del lenguaje**
- Plan de Impulso de las Tecnologías del Lenguaje
- Red nacional y europea de centros de excelencia en IA
- La sede principal en el campus de Gipuzkoa
 - Interdisciplinar: 7 departamentos de la Universidad



Universidad
del País Vasco

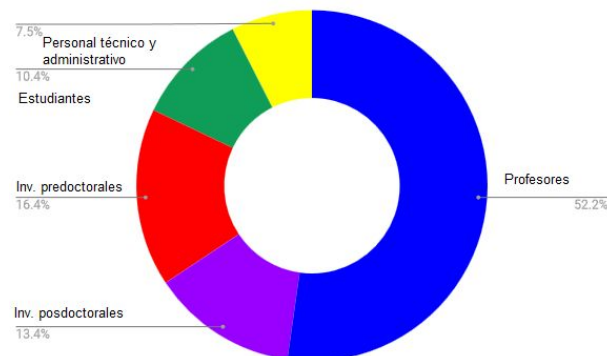
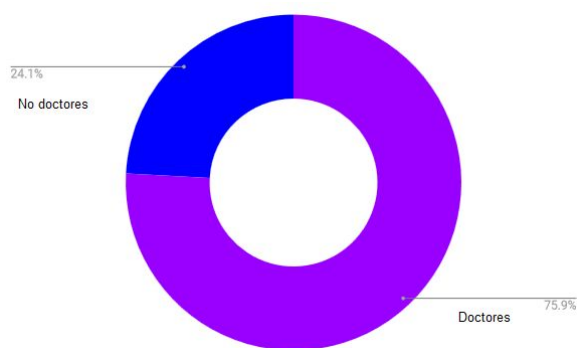
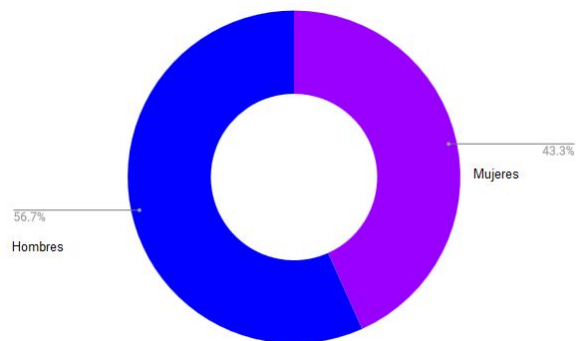
Euskal Herriko
Unibertsitatea

HiTZ

Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology



Personal: 67 personas



eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

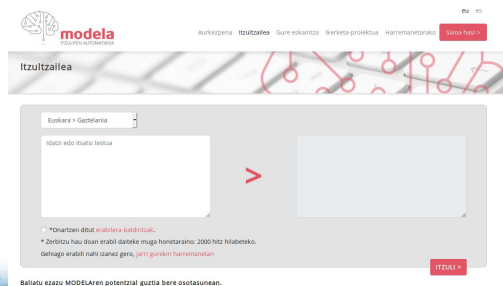
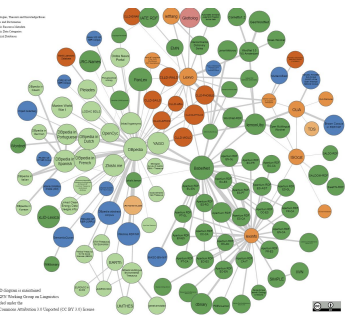
HiTZ

Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology



Patrimonio tecnológico del centro HiTZ

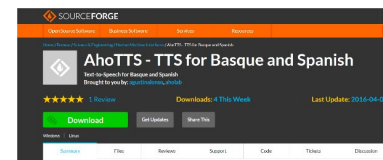
- **Recursos, herramientas básicas para voz y texto para lenguas diferentes:** euskara, castellano, inglés...
- **Traductores automáticos:**
 - Experiencia de más de 25 años: Matxin, SMT, MODELA
 - Actualmente: técnicas de Deep Learning con muy buenos resultados
- **Sistemas de extracción de información:** a partir de voz/texto y multilingüe
- **Sistemas de minería de textos en dominios y sectores específicos:** Justicia, Salud, Turismo, Enseñanza, Administración pública y sector privado.
- **Sistemas de inteligencia competitiva** en medios sociales y redes sociales, prensa digital, etc.
- **Sistemas de diálogos (chatbots)**
- **HPC - Computación de altas prestaciones**
 - Big Data
 - Deep Learning





Producción

- **Experiencia puntera** en el campo de las tecnologías lingüísticas
- Producción científica más de **60 artículos científicos al año** en revistas y congresos de reconocido prestigio
- Pertenencia a **redes de investigación** europeas y nacionales:
 - CLAIRE, EUROPEANA, Flarenet, TIMM, RTTH y RETELE
 - Promotor del centro CLARIN-K en España
- **Premios:** Google Research Award (2016 y 2019), Ayudas BBVA a equipos de investigación científica Big Data (2019), el premio Anton Abadia (2013) y el Manuel Laborde Werlinden (2001)
- **Transferencia tecnológica:**
 - Más de 30 empresas están explotando nuestros productos registrados
- **Más de 100 proyectos** en 30 años (1989-2019)





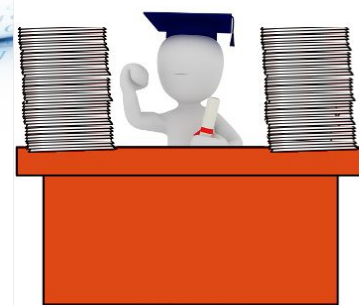
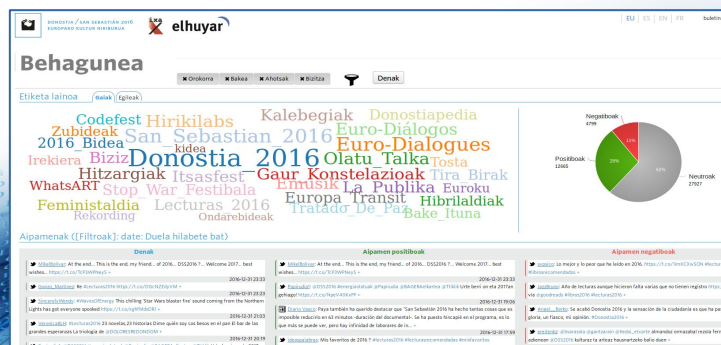
Formación

- Máster oficiales
 - Hizkuntzaren Azterketa eta Prozesamendua / Language Analysis and Processing
 - Erasmus Mundus Language and Communication Technologies
- Programa de doctorado
 - Análisis y Procesamiento del Lenguaje
 - 60 tesis leídas, 25 en progreso
- Cursos de perfeccionamiento
 - Deep Learning for Natural Language Processing





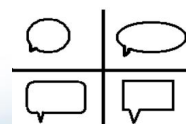
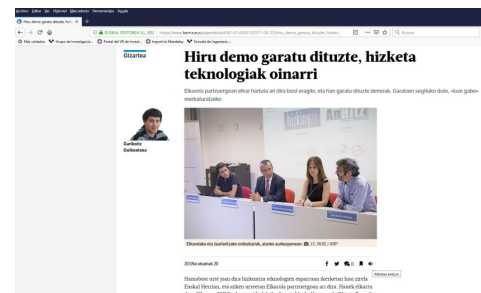
- 
- A photograph of two men standing in an office. The man on the left is wearing a dark hoodie, and the man on the right is wearing a dark t-shirt. They are both smiling at the camera. Behind them are shelves filled with books and papers, and a desk with a computer monitor and other office supplies. A window is visible in the background on the right.
- El Confidencial**
- Estos ingenieros vascos han creado un traductor para hacer sombra a ...**
- Traductor: Estos ingenieros vascos han creado un traductor para hacer sombra a Google, Noticias de Tecnología. Un equipo de investigadores de la Universi...
- elconfidencial.com





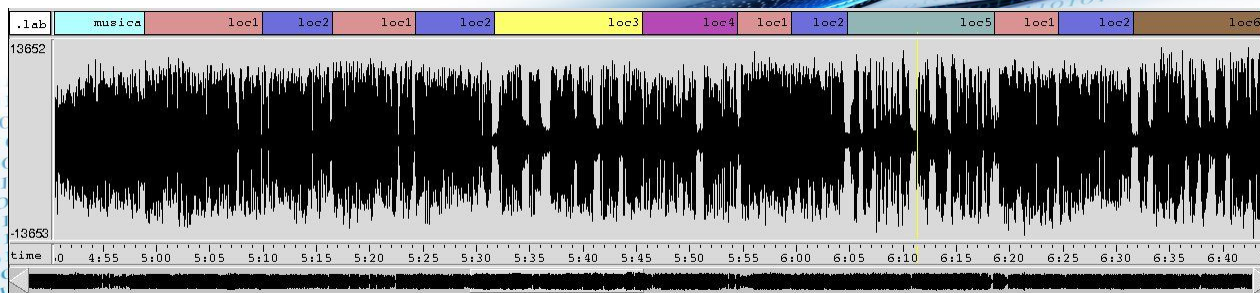
Líneas de investigación: procesamiento de voz

- Conversión de texto en habla
- Reconocimiento automático de voz
- Tecnologías del habla y diversidad funcional
- Extracción de información a partir de voz y audio: *Machine listening*
- Generación de recursos, herramientas básicas, y mejora de los existentes
- Tecnologías del habla aplicadas a la enseñanza



Ahospeaker: Speech Database in Basque for Synthesis and Voice Conversion

500 phonetically balanced sentences recorded by 4 Male and 3 Female voice talents for Neutral style. It was registered at 48kHz, 16bits, semi-professional room, 2 microphones and laryngograph included.



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

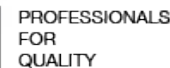
HiTZ

Hizkuntza Teknologiko Zentroa
Basque Center for Language Technology

Red de colaboradores



AITA MENNÍ



Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology



Miembros de HiTZ en la prensa en las últimas semanas

¿Hablar con un buscador de internet? Google premia a un investigador vasco para conseguirlo

El profesor de la UPV/EHU Eneko Agirre desarrollará su trabajo en inteligencia artificial para crear sistemas de diálogo que puedan conversar con las máquinas sobre la información existente en internet

INNOVADORES 07 MAY. 2019

5 minutos



Eneko Agirre (a la derecha) y colaboradores en el estudio. / Nagore Inaola - UPV/EHU

ITZIAR GONZÁLEZ-DIOS

Hizkuntzalari konputazionala

“Ez nekien non ari nintzen sartzen; hortaz, ez nion beldurrik”

Galegoa Aiztaran, Ama
Eneko Agirre



Itziar González-Dios. Itziar González-Dios lortu du hizkuntza teknologia eta komunikazio suloak indartzea da EHUk, eta kontrola jo ohi diren bi mundu batzen dituen bere jardueren: hizkuntzalari (Leta) eta informetika (Zientziak).

EUSKAL HERRIA

ERREPORTAJEA

FACEBOOK-EN BEKA BAT, IKERTZAILEEN MUNDU PREKARIZATUAN ARNESA

MADRID

INPRIMATU

BIDALI

Plazaberrri

Euskaltzaindia / Komunikazioa / Plazaberrri / Gero eta euskarazko doktoretza-tesi gehiagok lortzen dute nazioarteko kalifikazioa



Gero eta euskarazko doktoretza-tesi gehiagok lortzen dute nazioarteko kalifikazioa

2019-05-08

- 2018. urtean, 33 tesi defendatu ziren euskaraz UPV/EHU, eta horietatik 17 nazioartekoak izan ziren, aurreko urteko kopurua bikoiztuz. Datu hori nabarmendu dute UPV/EHUko eta Euskaltzaindiko ordezkariak, Euskarazko Tesien VI. Koldo Mitxelena sariak banatu diren egunean.
- Hauek izan dira 2018ko sarituak: Asier López (Zientziak); Olatz Pérez de Viñaspre (Ingeniaritza eta Arkitektura); Mikele Macho (Osasun Zientziak); Juan Mari Aguirreurreta (Gizarte eta Lege Zientziak); eta Miren Ibarluzea (Arte eta Giza Zientziak). Diploma bana eta 2.000 euro gordineko diru-kopurua jaso dituzte.

iatzak 8 dituela, Euskarazko Tesien VI. Koldo Mitxelena sariak banatu dira Euskaltzaindiaren Bilboko egoitzan. Sariok bost jakintzan-anatzen dira: Zientziak; Ingeniaritza eta Arkitektura; Osasun Zientziak; Gizarte eta Lege Zientziak, eta Arte eta Giza Zientziak. hurren, hauek izan dira sarituak: Asier López; Olatz Pérez de Viñaspre; Mikele Macho; Juan Mari Aguirreurreta, eta Miren . Diploma eta 2.000 euroko diru-kopurua jaso ditu bakoitzak.



IKERKETA

«Gure ekarpenek erabilera praktikoa izatea espero dugu, baina oinarritzko ikerketa ez da egiten epe motzeko aplikazio praktikoetan bakarrik pentsatuz»

Hizkuntzei zientziaren begietatik beha ari da Mikel Artetxe, eleaniztun egiten ordenagailu eta adimen artifizialak, agian, praktikan, euskara bezalako hizkuntza minorizatuendako mesedetan, baina batez ere zientziaren garapeneren izenean, esploratu ez diren espazioetara iristeko. 26 urte ditu Artetxek, Ingeniaritza Informatikoan graduatua da, Hizkuntza Azterketa eta Prozesamendua masterra egin du eta EHUko Informatika fakultateko IXA ikerketa taldearen baitan doktorego tesia egiten ari da orain, hirugarren urtez, gizakien gainbegiratzetik gabeko itzultzaile automatikoen inguruan.

dazioa > Sarean.eus

4. Irratsaioa. Inma Hernaez 2019/05/04

2 | 0 | 0



Descargar

Compartir

Me gusta

Más

Descripción de Sarean 64. Irratsaioa. Inma Hernaez 2019/05/04

Gaurko saioan ahotsaren inguruan hitz egingo dugu, Inma Hernaezekin. Ahotza nola sintetizatzen den eta zer den ikasiko dugu.

Jarraitzeko, Azkue Fundazioaren eskutik, Iurdana Acausorekin egongo gara zuzenean, Azkue Fundazioak ekarriko dizkigun ikastaro teknologiko berriak aurkezteko.

google ikastaroa Smart Compose ahots sintetizatua

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

HiTZ

Hizkuntza Teknologiko Zentroa
Basque Center for Language Technology

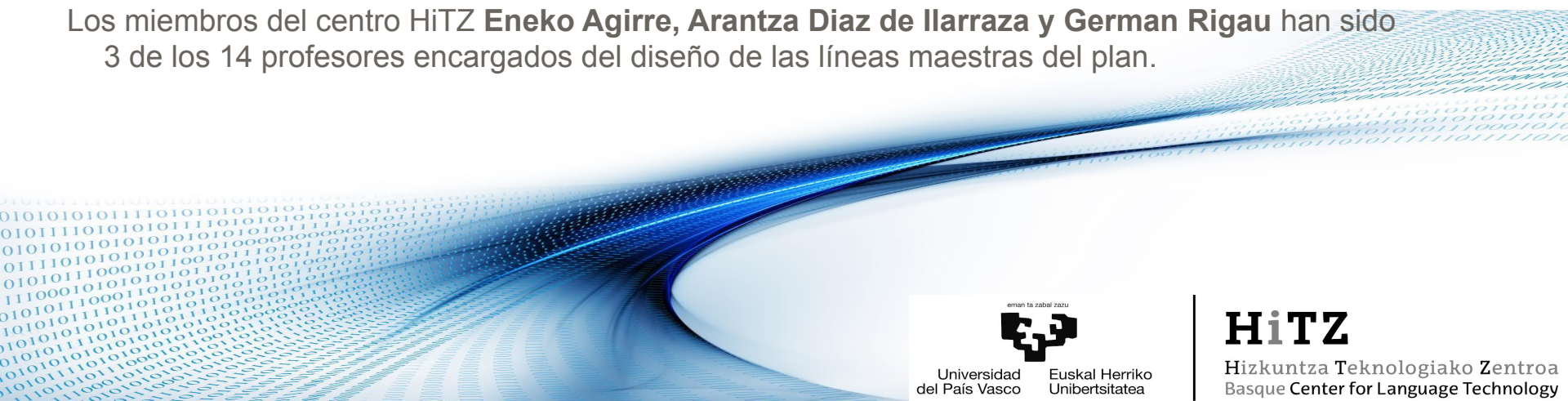


Plan de Impulso de tecnologías del lenguaje

- Creado en **2015** por la SESIAD (Ahora SEAD)
(Secretaría de Estado para la Sociedad de la Información y la Agenda Digital)
- **90 M€** para un periodo de 5 años.
- **Fomentar PLN, traducción automática y sistemas conversacionales** en lengua española y lenguas cooficiales
- **Objetivos:**
 - Mejora las **infraestructuras lingüísticas**
 - **Transferencia** de conocimiento: investigación <-> industria
 - **Internacionalización** de las empresas e instituciones
 - Generación, estandarización y difusión de **recursos lingüísticos sobre gestión pública** propia de la Administración



Los miembros del centro HiTZ **Eneko Agirre, Arantza Diaz de Ilarraza y German Rigau** han sido 3 de los 14 profesores encargados del diseño de las líneas maestras del plan.



Tecnologías del lenguaje en la frontera de la IA
Nuevo centro de referencia en tecnologías del lenguaje:

HiTZ

Hizkuntza **T**eknologiako **Z**entroa
Basque Centre for Language Technology



HiTZ

Hizkuntza **T**eknologiako **Z**entroa
Basque Center for Language Technology

Machine Listening for Car Sound Analysis

Joint Research Project by Mercedes-Benz Vitoria and Aholab Signal Processing Laboratory

Context

- The sound of the vehicle carries important information: Acoustic comfort and quality perception, assembly faults, mechanical failures,... but it is mixed with a very noisy background.
- Human perception of sound can be very precise but it is also highly subjective.
- Can a machine catch up the human hearing, or even exceed it, detecting and identifying undesirable noises?

Objective

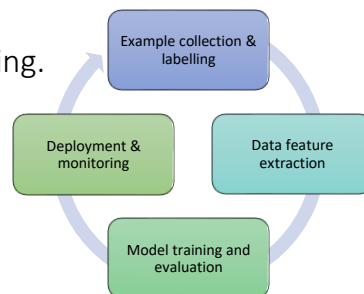
Create an automated tool to reproduce the sound assessment of the human auditors in real driving conditions.

- Detect undesirable noises (Buzz, Squeak & Rattle).
- Rate noise severity according to human auditors.
- Using the global sound field of the car interior.
- Driving in a real road and in a *Shaker*.



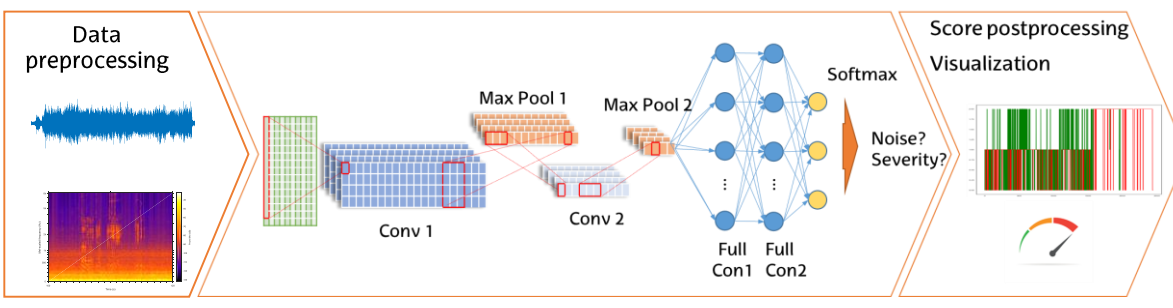
Methodology

- Recording of a sound event database and manual labelling.
- Data preprocessing and feature extraction.
- Build and train a Machine Learning system.
- Deployment and performance monitoring.
- Feedback and retraining.



Technology

- Deep Neural Networks are good at modelling complex processes and tackling high background noise levels.
- Convolutional NN to recognise the time-frequency patterns of the noises.
- Trained with examples from tens of vehicles and noise events.



Architecture



Capture equipment

- Up to 6 omnidirectional microphones with a recording device



User equipment

- Standard computer
- Processing of the recordings and feature extraction
- Queries to the inference server
- Visualization tools



Inference server

- Applies models to the inference queries
- Model retraining

Applications

- Sound quality assessment
- Assembly quality check
- Automatic failure diagnosis
- Predictive failure detection