Coordinated project: Voice Restauration with Silent Speech Interfaces **Acronym:** ReSSInt

Voice Restauration with EMG Silent Speech Interfaces (Bilbao)

Voice Restoration with Brain Computer Interfaces (Granada)

Silent Speech Interfaces

Allow humans to communicate by speaking silently

Use signals acquired from the human body while speaking silently

Strategies

Signals

Part of speech production process from which signals are extracted:
- Muscles of the face, tongue, jaws…
- Brain

Devices

Permanent Magnetic Articulography
Electromagnetic Articulography
Electromyography

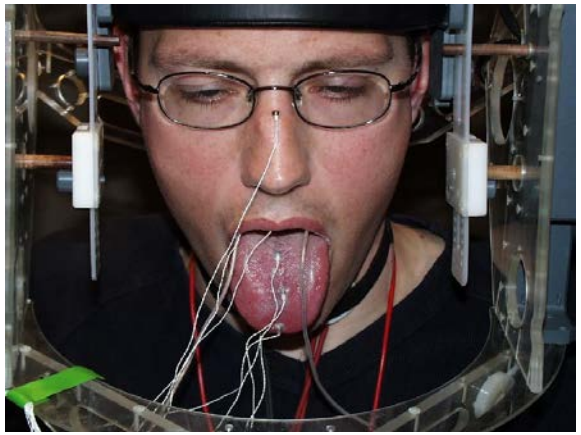
Electroencephalography
Electrocorticography
…

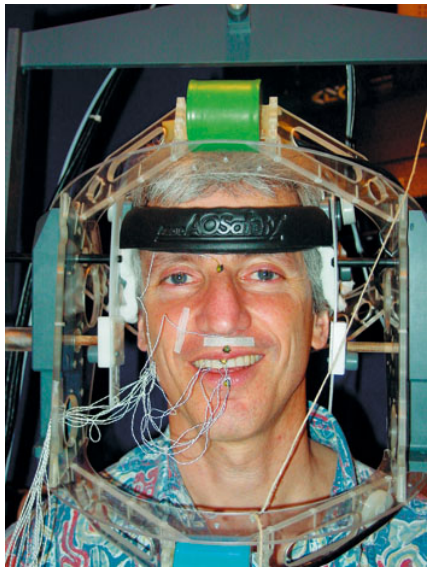
Format of Message

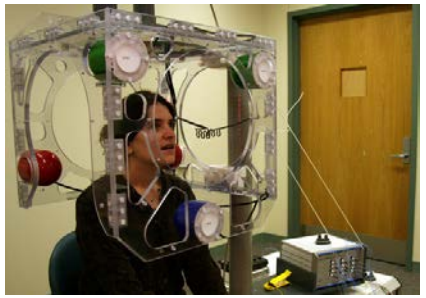
Text (speech recognition)
Speech (speech synthesis)

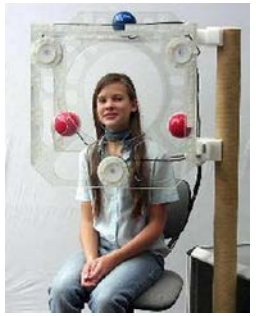
Silent Speech Interfaces

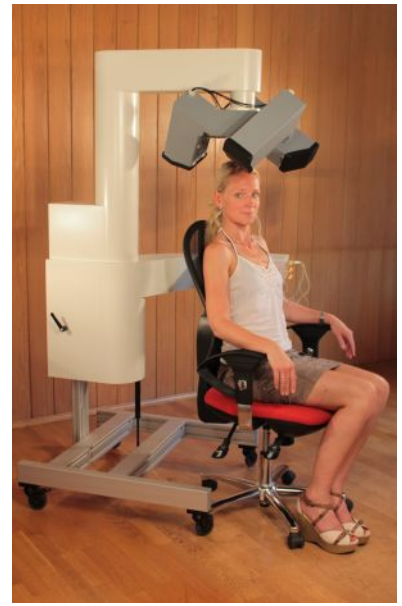
The Electromagnetic Articulograph





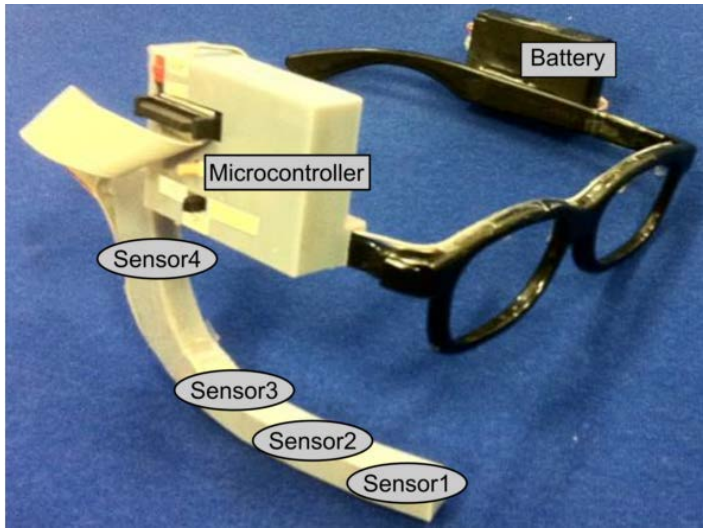


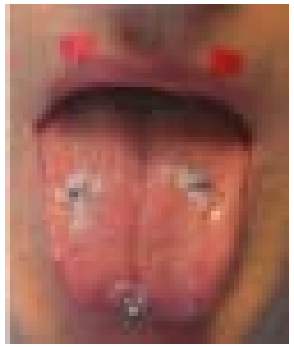


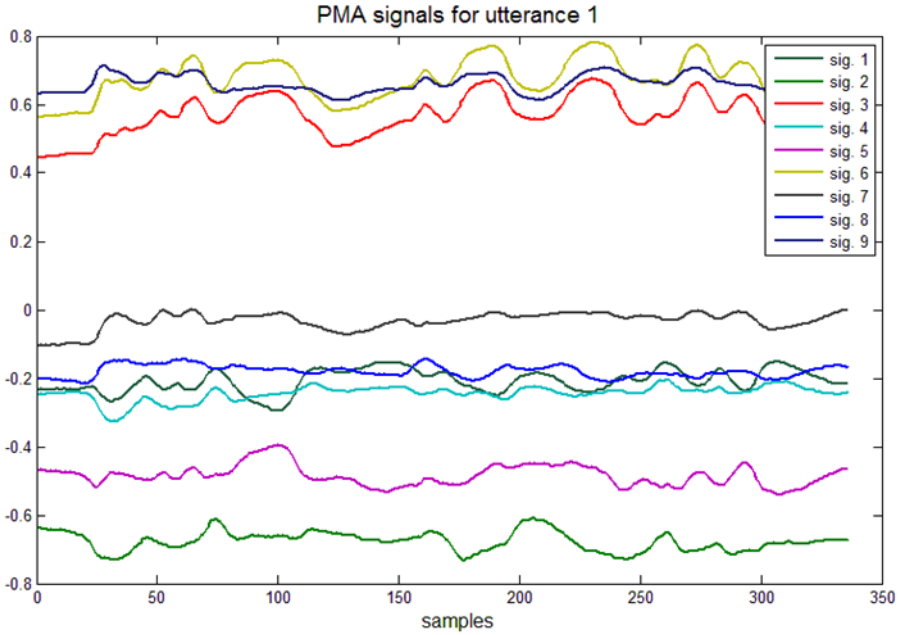


Silent Speech Interfaces

Permanent Magnet Articulography







Original

Convertida (F0 copiada)

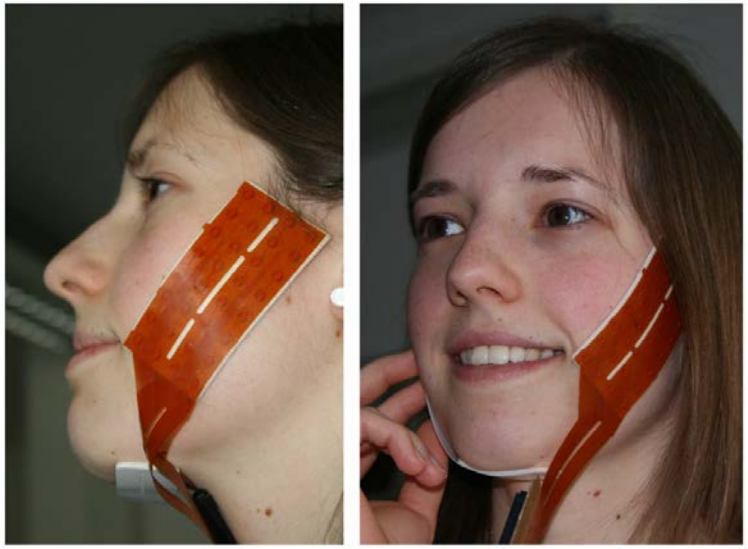
JA Gonzalez, LA Cheah, JM Gilbert, J Bai, SR Ell, PD Green, RK Moore. A Silent Speech System based on Permanent Magnet Articulography and Direct Synthesis. Computer Speech and Language *Vol 39, 2016, pp 67-87*.

Examples: PMA 2 Speech (F0 copied)

	Ex 1	Ex 2	Ex 3	Ex 4
Original				
Vocoded				
JDM				
WLR				
MLPG+GV				
MGE+GV				

Silent Speech Interfaces

Electromiography

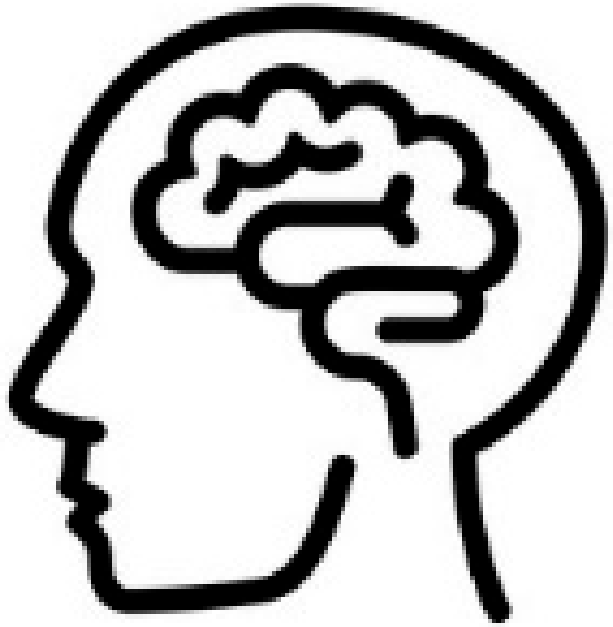


EMG-to-Speech: Direct Generation of Speech From Facial Electromyographic Signals (Matthias Janke, Lorenz Diener), In IEEE/ACM Transactions on Audio, Speech and Language Processing, volume 25, 2017.

Cognitive Systems Lab

https://www.uni-bremen.de/en/csl/research/silent-speech-communication/

Silent Speech Interfaces



Silent Speech Interfaces Applications

01

Silent telephony

Silent speech recognition allows for silent communication without disturbing any bystanders

02

Transmitting confidential information

Safely and securely transmission of confidential information such as passwords and PINs

03

Robust communication in adverse environments

Performs well in noisy and adverse conditions, because the processed signals are not acoustic.

04

Help for disabled people

Can help people who have lost their voice due to accident or illness

Team

SP1


Inma Hernáez


Eva Navas


Ibon Saratxaga


Jon Sanchez


Víctor García


Tanja Schultz


Lorenz Diener


Michael Wand

SP2


José A González


José L Pérez-Córdoba


Gonzalo Olivares


Alberto Galdón


Tanja Schultz


Phil D Green


Ricard Marxer

Objectives of the coordinated project

- IMPROVE QUALITY & INTELLIGIBILITY OF CURRENT SILENT SPEECH INTERFACES BASED ON EMG & ECoG
- DEVELOP CORPUS, DATABASES, PROTOCOLS & BEST PRACTICES
- ESTABLISH A RESEARCH INFRASTRUCTURE FOR SILENT SPEECH INTERFACES IN SPAIN
- STRENGTHEN LINKS BETWEEN AHOLAB (UPV/EHU) & SiGMAT (UGR)

Objectives of SP1

01

Establish infrastructure for the acquisition and processing of EMG signals

02

Develop a high-quality baseline EMG-based direct speech synthesis system using DNNs, including the necessary databases

03

Investigate novel architectures to overcome the problem of inter-session and inter-speaker variability

04

Validate the use of EMG SSI to be used by laryngectomees

Objectives of SP2

01

Record the first large-scale data corpus in Spanish with (a) parallel speech & intracranial neural recordings and (b) non-parallel recordings for imagined speech with only brain signals

02

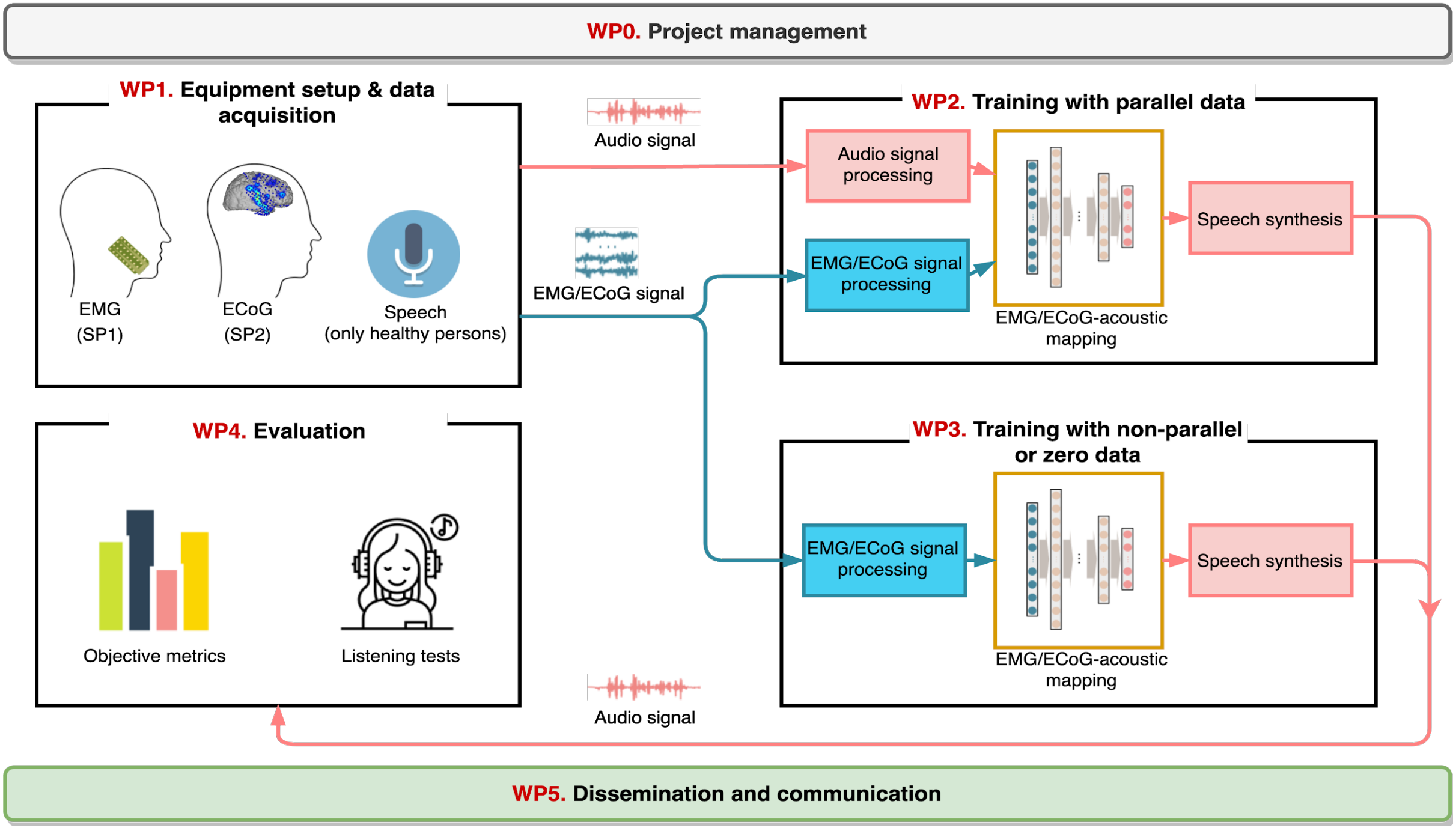
Develop a high-quality baseline ECoG-to-speech system trained with parallel data

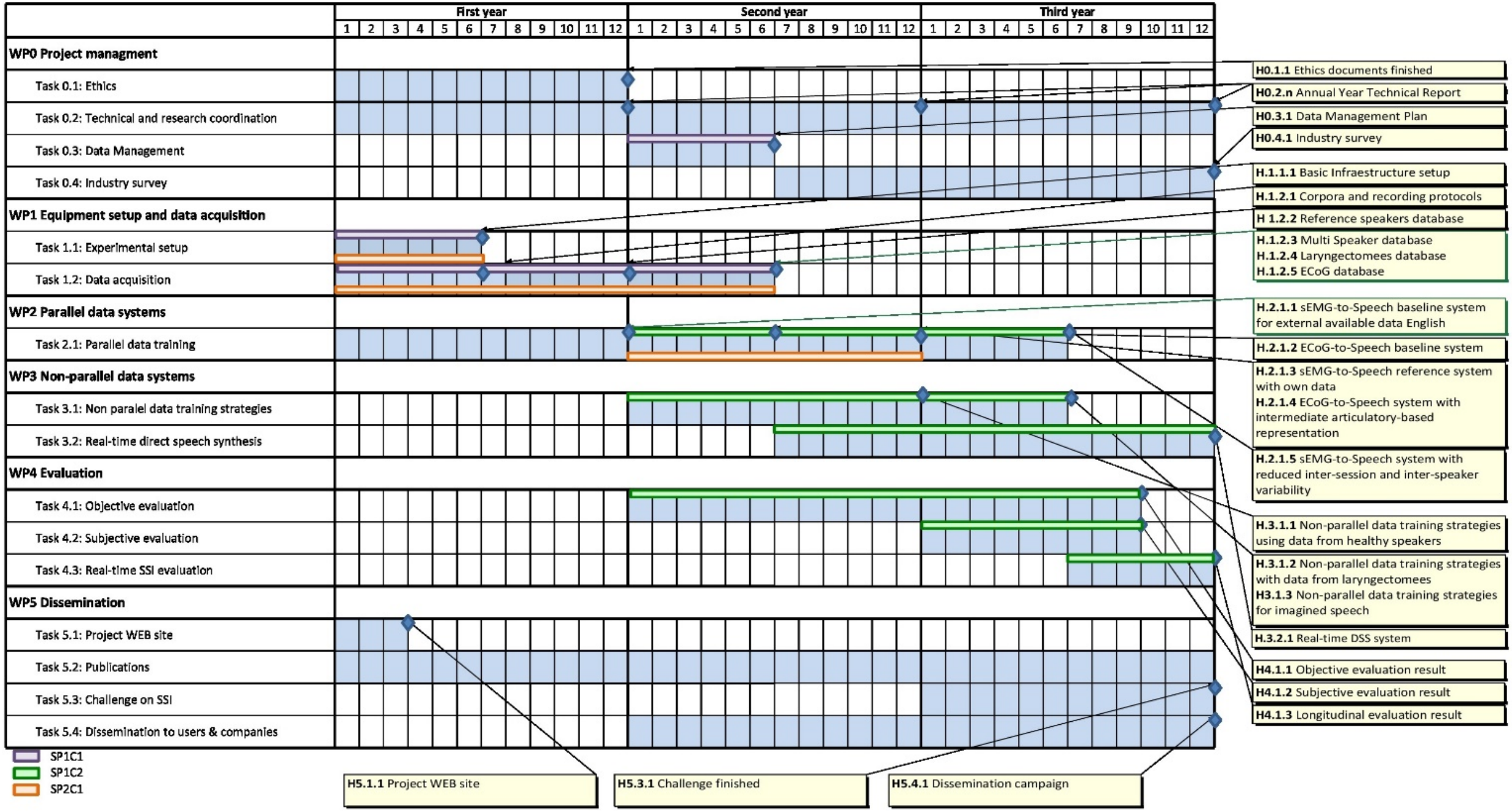
03

Investigate the use of transfer learning to adapt pre-trained DNN models trained on parallel data for the task of synthesize imagined speech

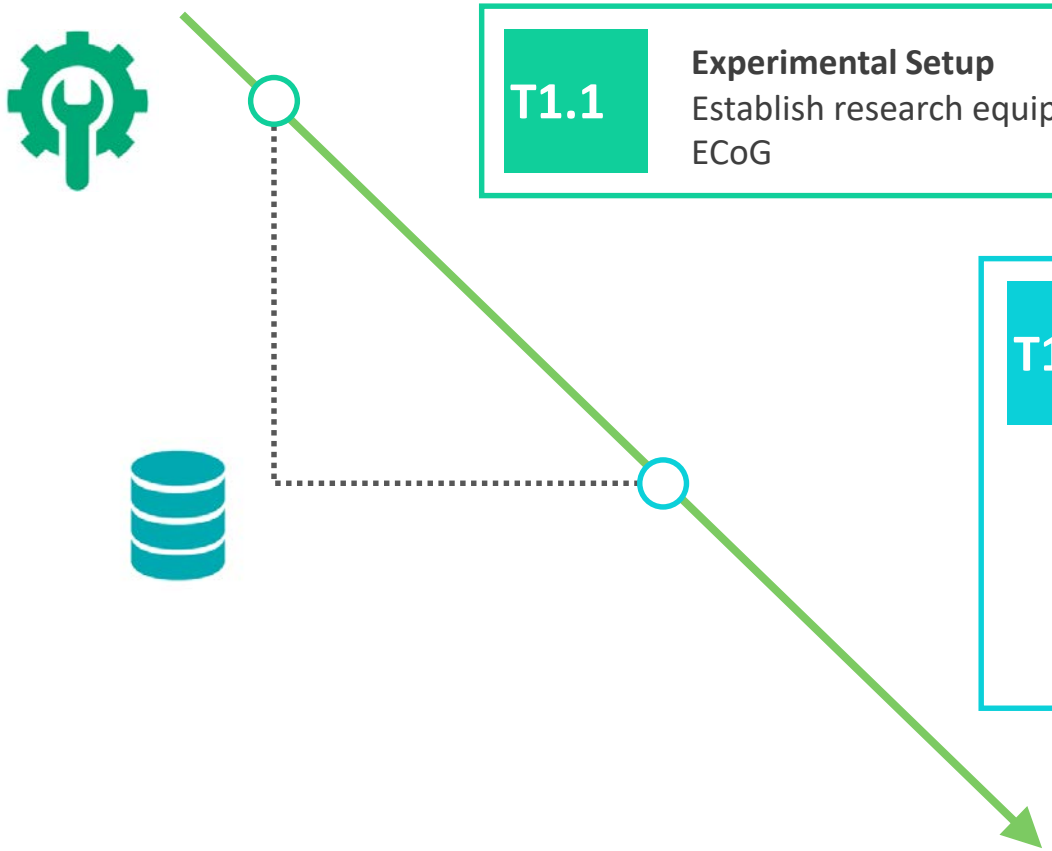
04

Investigate novel algorithms for DNN training with non-parallel data for direct speech synthesis from imagined speech


WP0. Project management
WP1. Equipment setup & data acquisition
EMG
(SP1)
ECoG
(SP2)
Speech
(only healthy persons)
Audio signal
EMG/ECoG signal
WP2. Training with parallel data
Audio signal processing
EMG/ECoG signal processing
EMG/ECoG-acoustic mapping
Speech synthesis
WP3. Training with non-parallel or zero data
EMG/ECoG signal processing
EMG/ECoG-acoustic mapping
Speech synthesis
WP4. Evaluation
Objective metrics
Listening tests
Audio signal
WP5. Dissemination and communication


First year
Second year
Third year
1 2 3 4 5 6 7 8 9 10 11 12
WP0 Project managment
Task 0.1: Ethics
Task 0.2: Technical and research coordination
Task 0.3: Data Management
Task 0.4: Industry survey
WP1 Equipment setup and data acquisition
Task 1.1: Experimental setup
Task 1.2: Data acquisition
WP2 Parallel data systems
Task 2.1: Parallel data training
WP3 Non-parallel data systems
Task 3.1: Non paralel data training strategies
Task 3.2: Real-time direct speech synthesis
WP4 Evaluation
Task 4.1: Objective evaluation
Task 4.2: Subjective evaluation
Task 4.3: Real-time SSI evaluation
WP5 Dissemination
Task 5.1: Project WEB site
Task 5.2: Publications
Task 5.3: Challenge on SSI
Task 5.4: Dissemination to users & companies
SP1C1
SP1C2
SP2C1
H0.1.1 Ethics documents finished
H0.2.n Annual Year Technical Report
H0.3.1 Data Management Plan
H0.4.1 Industry survey
H.1.1.1 Basic Infraestructure setup
H.1.2.1 Corpora and recording protocols
H 1.2.2 Reference speakers database
H.1.2.3 Multi Speaker database
H.1.2.4 Laryngectomees database
H.1.2.5 ECoG database
H.2.1.1 sEMG-to-Speech baseline system for external available data English
H.2.1.2 ECoG-to-Speech baseline system
H.2.1.3 sEMG-to-Speech reference system with own data
H.2.1.4 ECoG-to-Speech system with intermediate articulatory-based representation
H.2.1.5 sEMG-to-Speech system with reduced inter-session and inter-speaker variability
H.3.1.1 Non-parallel data training strategies using data from healthy speakers
H.3.1.2 Non-parallel data training strategies with data from laryngectomees
H3.1.3 Non-parallel data training strategies for imagined speech
H.3.2.1 Real-time DSS system
H4.1.1 Objective evaluation result
H4.1.2 Subjective evaluation result
H4.1.3 Longitudinal evaluation result
H5.1.1 Project WEB site
H5.3.1 Challenge finished
H5.4.1 Dissemination campaign

WP1 Experimental setup & data acquisition



T1.1

Experimental Setup
Establish research equipment setup both for EMG & ECoG

T1.2

Data acquisition
Provide training & testing data both for EMG & ECoG

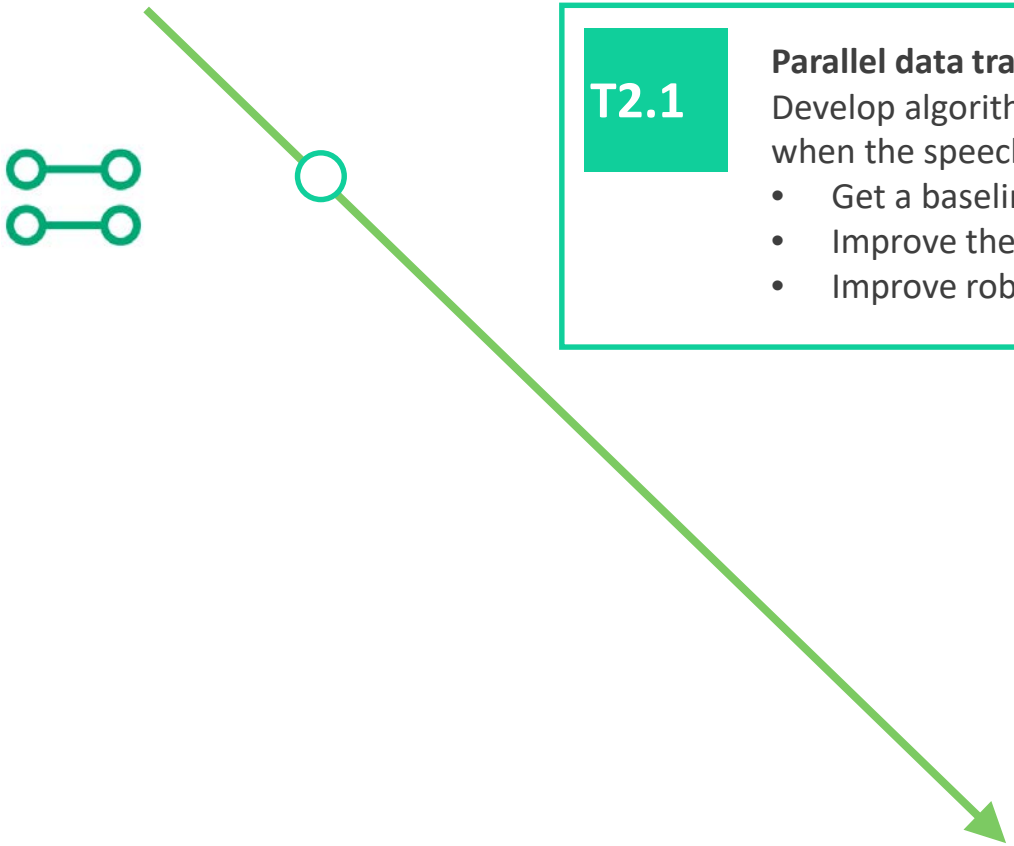
SP1:

- Corpus of small number of healthy subjects, several sessions, with speech. Some with many sessions
- Corpus of laryngectomized subjects

SP2:

- ECoG-Speech parallel corpus and ECoG only corpus

WP2 Parallel data systems



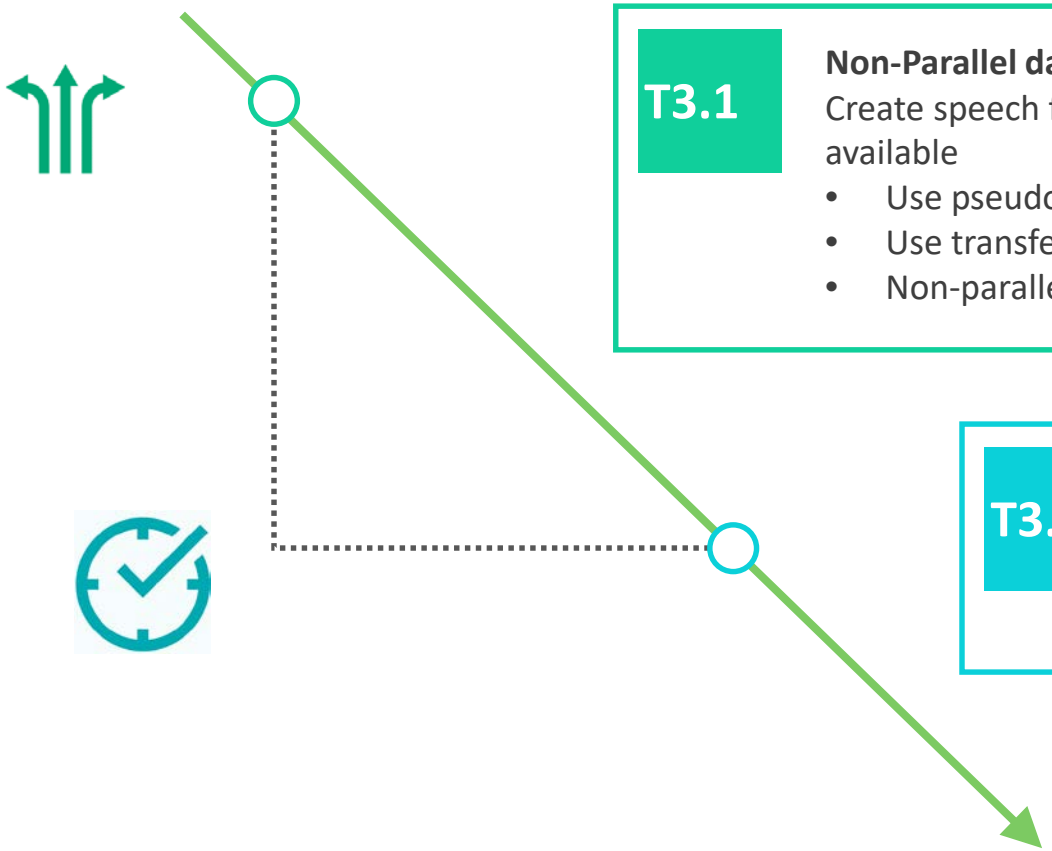
T2.1

Parallel data training strategies

Develop algorithms to generate speech from biosignals when the speech signal has also been recorded

- Get a baseline system with data previously available
- Improve the speech produced
- Improve robustness to speaker & session variability

WP3 Training strategies for non-parallel data or zero-data training



T3.1

Non-Parallel data training strategies

Create speech from biosignals when no audible speech is available

- Use pseudo-parallel data (mimicking, synthesis)
- Use transfer learning
- Non-parallel voice conversion (CycleGAN)

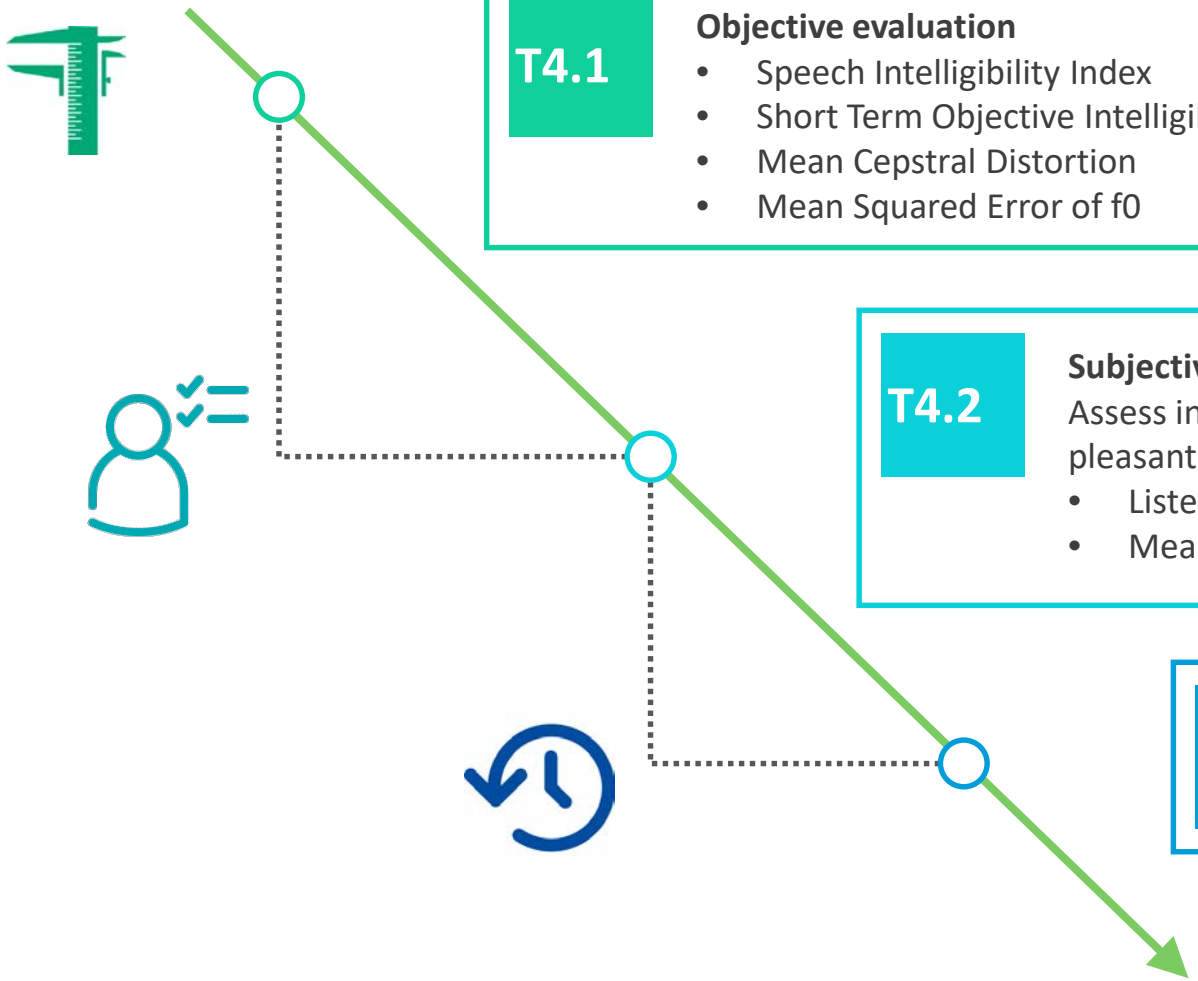
T3.2

Real time Direct Speech Synthesis

Optimize the system to get very low latency

- User can adapt to the system and improve the results
- System can be adapted to the particular user

WP4 Evaluation



T4.1

Objective evaluation

- Speech Intelligibility Index
- Short Term Objective Intelligibility
- Mean Cepstral Distortion
- Mean Squared Error of f0

T4.2

Subjective evaluation

Assess intelligibility, naturalness and pleasantness of the generated speech signals

- Listen and transcribe task
- Mean Opinion Score

T4.3

Evaluation along time

Evaluate the ability of users to adapt to the SSI system

Questions