

Non-native Speech in European Portuguese

1. BASIC INFORMATION

1. *Corpus composition*

This corpus is composed of audio of native and non-native speakers of European Speakers. It is comprised of read and spontaneous speech. The 'speakers_information.txt' file gives information about the speakers (age, gender, childhood district). Nevertheless, this information is not available for some speakers.

2. *Representation of the corpora (flat files, database, markup)*

The corpus is comprised of flat audio files organized in directories: natives/non-natives, read/spontaneous speech, and per speaker.

2. ADMINISTRATIVE INFORMATION

1. *Contact person*

Name: Isabel Trancoso
Address: Rua Alves Redol, nº 9, 1000-029, Lisboa
Affiliation: IST/INESC-ID
Position: Full Professor
Telephone: +351-213-100-268
Fax: +351-213-145-843
e-mail: isabel.trancoso@inesc-id.pt

2.2 *Delivery medium (if relevant; description of the content of each piece of medium)*

The resource will be uploaded on the MetaShare platform as a ZIP archive.

2.3 *Copyright statement and information on IPR*

The resource is distributed under the CC-BY licence.

3. TECHNICAL INFORMATION

1. *Directories and files*

The archive that will be uploaded on the MetaShare platform will contain two folders separating native and not native speakers, each of which is organised by speech type (read and spontaneous), and lastly by speaker, containing the audio .wav files.

2. *Corpora size (nmb. of tokens, MB occupied on disk)*

The whole resource occupies 713 MB. The total time of all the audio files is 213 min (3h33min).

4. CONTENT INFORMATION

1. *Type of the corpus*

This corpus is monolingual and is only comprised of speech files.

2. *The natural language(s) of the corpus*

The language of the corpus is European Portuguese, with native and non-native speech. The non-native speakers come from a variety of countries.

5. *Intended application of the corpus*

Studies on non-native speech in European Portuguese