## Phonetic Events from the Labelling of the European Portuguese Database for Speech Synthesis, FEUP/IPB-DB

João Paulo Teixeira<sup>1</sup>, Diamantino Freitar<sup>2</sup>, Daniela Braga<sup>2</sup>, Maria João Barros<sup>2</sup>, Vagner Latsch<sup>2</sup> <sup>1</sup>EST63 – Instituto Politécnico de Bragança <sup>2</sup> Faculdade de Engenharia da Universidade do Porto joaop€ajb.b.p. dfreitasse<sup>8</sup>€e.up.p.t. dbrazaga<sup>8</sup>€e.up.p.t. mbarrose<sup>8</sup>€e.up.p.t. vagners<sup>8</sup>€e.up.p.t



We present a new labelled speech signal database (FEU/PIPB-DB) in Standard European Portuguese. The FEU/PIPB-DB provides phonetic material for construction of Text-to-Speech (TTS) systems and puts at SEP scientific community's disposal a phonetically and prosodically valuable speech corpus, essential for Prosody modelling and Phonetic research.

Some methodological problems are discussed and some observed phenomena of experimental phonetics deriving from the speech signal labeling are presented with some resulting statistics of segments' lengths. The multidisciplinary team involved in the project provides the phonetic, linguistic and technical consistency.

### Portuguese Speech DB

· Some\_examples of Labelled Tracks are available at:

- http://www.portugues.mct.pt/Repositorio/EuroSpeechIPB/
- · A total of 100 minutes of recorded material.
- A professional male speaker read some text materials in an acoustically treated professional studio. Speech was digitally recorded at 44.1 KHz, 16 bits, mono.
- · Text Corpus newspapers articles (see detail on the right).
- Labelling was manually done by a professional phonetician using the Speech Filing System tool for almost all the materials.
- · Labelling (see detail on the right):
- Segments of Phonemes (Basic SAMPA code + Silence + tonic syllable)
- · Words (beginning and end of each word)
- · Phrases (beginning and end of each phrase + punctuation marks)



Some phonetic events reported by the Phonetician responsible for the manual labelling of the DB:

•Dialectal Changing

|  | Example   | Standard EP | Dialectal change            |  |  |  |
|--|-----------|-------------|-----------------------------|--|--|--|
|  | doutores  | /o/         | /ow/ diphtongization        |  |  |  |
|  | hoje      | /o/         | /oj/, /je/ diphthongization |  |  |  |
|  | elc       | /o/         | before palatal consonant    |  |  |  |
|  | regressou | /R/         | /r/ multiple alveolar trill |  |  |  |
|  | embora    | /e-/        | /6-i/                       |  |  |  |

#### Contextual Changing

| <ul> <li>Suppressions</li> </ul> | Suppressions     | [0]                       | [u]                       |
|----------------------------------|------------------|---------------------------|---------------------------|
| or Reductions                    | In the beginning | <explorado> -</explorado> | Not/A vailable            |
| of Reductions                    | In the middle    | <decisão> -</decisão>     | <português> -</português> |
|                                  |                  | [dsi"26~w]                | [prt"geS]                 |
|                                  | In the end       | <deve> - ["dEv]</deve>    | <porto> - ["port]</porto> |

 Vowel Quality Transformations (when two different vowels get together in an utterance)

- The vowels melt and experience a quality change (e.g. <fica admirado> [fikadmiradu]; <contra o> [kõtrO]).
- •One of the vowels, [@] or [i], becomes a semi-vowel (e. g. <se aprende> [sj6pr~ed]; <na idade> [n6jdad]).
- ·Additions ('schwas' between relaxed consonantal groups)

(e.g. <branco> [b@r6āku]).

•Allophones (phones changing motivated by the articulatory context)

-<-te > in a word final position followed by a pause→ [ts]

(the vowel [@] is reduced and the plosive 'fricatizes' with the closer articulatory point consonant [s]).

- -r > in a word final position followed by a pause
- (the [r] is longer and usually voiceless).
- •<l> in closed syllables is acoustically distinctive ( it was labelled with [1\*]).

• 'Fricatization' of voiced plosives in an inter-vocalic context  $(<-b \rightarrow \beta; <-d \rightarrow \delta; <-g \rightarrow \gamma)$ .

#### ·Statistical Information about the Phonemes

Some statistical information was extracted from 21 min. of speech that consists of about 18,700 phone segments at a speech rate of 12.2 phonemes/second. This information consists of occurrences frequencies (in %) and average durations of all phone segments, considering general and tonic syllable position (see detail on the right).

EUROSPEECH 2001 SCANDINAVIA

# Phonetic Events from the Labelling of the European Portuguese Database for Speech Synthesis, FEUP/IPB-DB

# Details)

### Portuguese Speech DB

•Text Corpus - articles from a Portuguese newspaper:

- · 11 text excerpts from different articles.
- · 1 interview.

 2 sets of specially prepared interrogative sentences, with and without interrogative pronouns (who – "quem", which – "qual", how many – "quantos", how – "como", where – "onde", etc.),

 1 set of phonetically engineered log-atoms carrying all standard Portuguese diphones and several triphones in a congruent context.

. Labels used in the phonetic, word and phrase levels:

| Code  | M caning                              |  |  |  |
|---|---------------------------------------|--|--|--|
| p, b, t, d, k, g  | Burst segment of plosive              |  |  |  |
|   | consonants in SAMPA code              |  |  |  |
| 1   | Stop segment of plosive               |  |  |  |
|   | consonants                            |  |  |  |
| f, v, s, z, S, Z  | Fricatives in SAMPA code              |  |  |  |
| m, n, J   | Nasals in SAMPA code                  |  |  |  |
| L, I, R, r  | Liquids in SAM PA code                |  |  |  |
| 1.  | l at the end of syllable (velar<br>l) |  |  |  |
| i, e, E, a, 6, O,<br>o, u, @  | Vowels in SAMPA code                  |  |  |  |
| $i \sim , c \sim , 6 \sim , o \sim , u \sim ,$<br>$w \sim , j \sim$ | Nasal vowels in SAMPA<br>code         |  |  |  |
| w, j  | Semi-vowels in SAMPA code             |  |  |  |
| x   | Silence                               |  |  |  |
| XX  | Aspiration                            |  |  |  |
|   | Beginning of tonic syllable           |  |  |  |
| Word level  |                                       |  |  |  |
| i   | Beginning of word                     |  |  |  |
| ſ   | End of word                           |  |  |  |
| Phrase level  |                                       |  |  |  |
| i   | Beginning of phrase                   |  |  |  |
| . f   | End of phrase                         |  |  |  |
| ,!?()-;:"   | All punctuation marks that            |  |  |  |

Phonetic, word and phrase labels from the speech excerpt presented on the left poster :

### « É a "guerra" declarada »

Timing Phoneme Label Timing Word Label Timing Phrase Label

|          | 33.37905 XX | 33.37995 f | 33.38095 . |
|----------|-------------|------------|------------|
| Ē        | 33.96020 E  | 33.96145 i | 33.96170 i |
| -        | 34.03925.6  | 34.03705 i |            |
| a        | 34.12205 1  | 34.12215 i | 34.12105 " |
| <b>A</b> | 34.17805 "g |            |            |
| ¥ .      | 34.19845 E  |            |            |
| u        | 34,34805 R  |            |            |
| ¥ .      | 34.43055.6  |            |            |
| ¥        | 34,491151   | 34.49325 i | 34.49035 " |
| a.       | 34.54150 d  |            |            |
| π.       | 34.55620 @  |            |            |
| d -      | 34,58030 1  |            |            |
| ž.       | 34.63870 k  |            |            |
| č        | 34.667651   |            |            |
| т        | 34.70830.6  |            |            |
| å.       | 34.78045 "r |            |            |
| F.       | 34.79650 a  |            |            |
| a        | 34.97040 1  |            |            |
| d i      | 35.03195 d  |            |            |
| ā        | 35.05660.6  |            |            |
| _        |             |            |            |

## **Phonetic Events**

•Statistical Information about the Phonemes

|           |      |      |     | Position |      |     |
|-----------|------|------|-----|----------|------|-----|
|           | 36   | Ay,  | xtd | %        | Ay,  | std |
|           |      | (ms) |     |          | (ms) |     |
| 3         | 4.0  | 110  | 34  | 2.2      | 121  | 32  |
| 6         | 10.0 | 68   | 28  | 0.8      | 75   | 33  |
| E         | 1.7  | 97   | 29  | 1.0      | 102  | 27  |
|           | 1.8  | 95   | 40  | 1.0      | 102  | 38  |
| 0         | 1.7  | 53   | 38  | 0.03     | 33   | 15  |
| 1         | 5.2  | 69   | 28  | 1.5      | 85   | 28  |
| 0         | 1.4  | 106  | 33  | 0.8      | 116  | 29  |
|           | 1.6  | 97   | 34  | 0.9      | 103  | 34  |
|           | 5.1  | 57   | 29  | 0.7      | 65   | 33  |
| 1         | 2.8  | 49   | 26  | 0.8      | 53   | 24  |
| w         | 2.5  | 44   | 27  | 0.7      | 47   | 31  |
| j-        | 0.1  | 64   | 20  | 0.03     | 61   | 21  |
| w-        | 0.04 | 53   | 31  | 0.02     | 55   | 34  |
| 6-        | 2.9  | 75   | 35  | 0.9      | 97   | 38  |
| 6-        | 1.2  | 107  | 31  | 0.6      | 117  | 33  |
| j         | 0.7  | 109  | 42  | 0.2      | 132  | 49  |
| 0-        | 0.9  | 98   | 36  | 0.3      | 119  | 41  |
| <b>D-</b> | 0.6  | 86   | 45  | 0.2      | 22   | 43  |
|           | 3.3  | 20   | 9   | 1.0      | 18   | 6   |
| -         | 3.3  | 64   | 19  | 1.0      | 70   | 19  |
| i         | 5.3  | 29   | 19  | 1.3      | 23   | 10  |
|           | 5.3  | 48   | 20  | 1.3      | 49   | 20  |
| k         | 41   | 37   | 16  | 1.0      | 36   | 11  |
|           | 41   | 59   | 17  | 1.0      | 61   | 16  |
| b         | 1.3  | 17   | 18  | 0.5      | 15   | 7   |
|           | 1.3  | 43   | 16  | 0.5      | 44   | 15  |
| d         | 4.7  | 20   | 17  | 0.8      | 15   | 5   |
|           | 4.7  | 41   | 17  | 0.8      | 39   | 15  |
|           | 1.3  | 20   | 13  | 0.6      | 19   | 7   |
|           | 1.3  | 44   | 13  | 0.6      | 43   | 12  |
| m         | 2.8  | 62   | 19  | 0.7      | 63   | 19  |
|           | 20   | - 54 | 19  | 0.4      | 51   | 15  |
| J         | 0.4  | 68   | 18  | 0.1      | 67   | 19  |
|           | 1.8  | 52   | 20  | 0.4      | 53   | 20  |
| 1*        | 0.9  | 68   | 30  | 0.4      | 78   | 32  |
| L         | 0.4  | 56   | 21  | 0.1      | 43   | 15  |
| r         | 6.5  | 32   | 16  | 2.1      | 34   | 17  |
| R         | 0.7  | 73   | 21  | 0.1      | 78   | 20  |
| v         | 1.4  | 65   | 22  | 0.3      | 69   | 20  |
| ſ         | 1.2  | 93   | 27  | 0.4      | 99   | 25  |
| z         | 1.6  | 70   | 18  | 0.4      | 74   | 19  |
| 8         | 4.2  | 103  | 31  | 1.1      | 100  | 28  |
| s         | 4.1  | 89   | 33  | 0.6      | 83   | 26  |
| Z         | 1.9  | 78   | 25  | 0.4      | 79   | 25  |
| XX        | 24   | 320  | 173 |          |      |     |
| x         | 2.6  | 165  | 210 |          |      |     |

It seems that only the vowels' and [1\*]' durations are affected by the tonic position.

## Conclusions

The presented DB is useful for prosodic analysis. Other type of prosodic information can automatically be derived from the DB with appropriate software. F0 can be extracted from the speech signal. Linguistic information, also useful for prosodic modelling, can be derived from original text material.

A set of example observations and results are presented and explained, at the statistical level, by frequencies of occurrence and basic segmental advantoms, and a the level of phonetic changing phenomena, by suppressions or reductions, vowel quality transformations, additions, allophones and phonetic changes. Although these aspects have a partial contribution to speech naturalness, they are essential for the segmental quality of a TTS system.

This DB has been used to build a diphone concatenative TTS for Portuguese.