

AMADEO: Apprentissage à partir de grandes masses de données orales

ABSTRACT

The AMADEO (*Apprentissage à partir de grandes masses de données orales*) project addresses the following scientific issues in speech processing: **how to improve the acoustic modeling of speech, by taking advantage of large annotated speech corpora?**

Word pronunciations vary as a function of various parameters (speaker, speaking style, accents, communication context...) Part of the observed variation can be considered as noise, part of it is contributing to the linguistic content.

At present acoustic pronunciation modeling does not consider information related to the **utterance modality** (e.g. affirmation/question) and **focus** (e.g. what is the utterance about). This information is mainly borne by prosodic features, which are also supposed to contribute to set boundaries and to follow potential modifications of the discourse topic. Prosody may also impact the **acoustic-phonetic realization** of speech. Automatic learning techniques applied to prosodic features together with syntactic/semantic tags are likely to contribute to progress in acoustic modeling of speech.

1 What is the AMADEO project ?

The AMADEO project provides a PhD funding on the subject of **learning with large corpora** in the domain of automatic speech and language processing, by combining the skills of two Digiteo labs: LIMSI/CNRS and CEA LIST.

The long term aim of the proposed research is to progress towards a more natural and intuitive oral human-machine interaction. Within the framework of AMADEO, the research objective of the ongoing PhD aims at improving prosodic and syntactic-semantic modeling of speech, by taking advantage of very large annotated speech corpora.

The following examples illustrate information from prosody and its link with focus, syntax and sense:

le départ de M. Chirac vs. le départ de M. Chirac (the departure of Mr. Chirac)

Focus on "départ (departure)" vs. on "M. Chirac (Mr. Chirac)".

à trois heures vingt # deux # place(s) non fumeur (at 3:22, 1 seat vs at 3:20, 2 seats).

Proposed studies aim at enhancing information extracted from audio signals to improve the possibilities of information retrieval systems and make Man-Machine interfaces by voice more attractive and intuitive. Such interfaces will have an important economic and societal impact.

Large **broadcast and conversational corpora** in French (several hundreds of hours) to be used for : (i) prosodic feature extraction and annotation, (ii) chunking and syntactic/semantic tagging of audio transcribed corpora, (iii) acoustic, syntactic/semantic and prosodic model estimation, (iv) reference tagged corpora creation for evaluation.

2 Investigated works

Can homophone words, bearing different POS tags, be distinguished using specific acoustic and prosodic parameters ?

→ acoustic measures - automatic classification using data mining techniques.

2.1 Large Corpus study

Corpora:

ESTER (Evaluation des Systèmes de Transcription enrichie d'Emissions Radiophoniques / prepared speech / 55h)

PFC (Phonologie du Français Contemporain / conversational speech / 12h).

LIMSI speech recognition system :

4-gram language models (LM) and context-dependent acoustic phone models.

best results for the 2005 ESTER evaluation: 11.9% WER (word error rate).

20 most frequent words contribute to 25% of observed word errors.

Selection of homophone words, bearing different POS tags :

et "and" vs. *est* "is"

à "to" vs. *a* "has" (very frequent in French and frequently misrecognized).

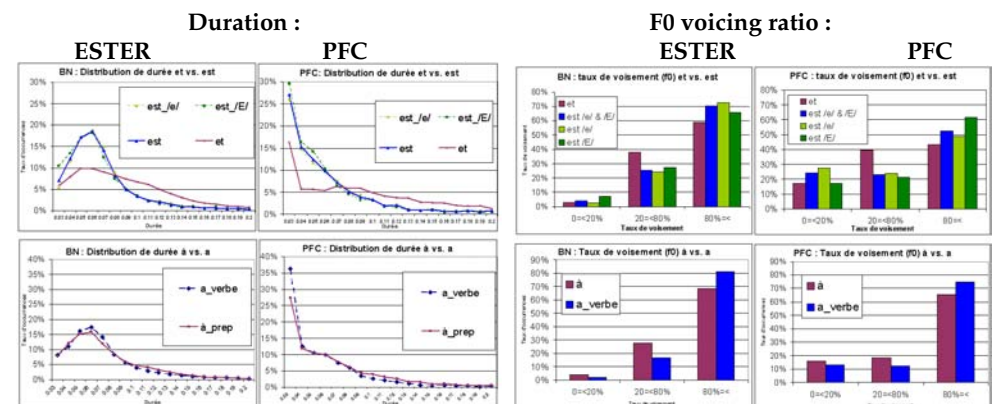
Occurrences:

pronunciation variants for *est*: /ɛ/ (canonical pron), /e/ (variant)

word	ESTER		PFC	
	occ.	/phoneme/	occ.	/phoneme/
à	20,4k	/a/	3,6k	/a/
a	11,3k	/a/	3,4k	/a/
et	19,1k	/e/	5,0k	/e/
est	14,5k	/ɛ/5,0k, /e/9,5k	6,2k	/ɛ/1,9k, /e/4,3k

2.2 Acoustic-prosodic parameter analysis

Duration, fundamental frequency (f0) voicing ratio, & pauses.



Different curves for *et* vs. *est* pair.

Impact of speaking style

Higher rates for POS=verb (*est* & *a*)

Pauses : silence, breath, hesitation.

Words	et		est		à		a	
	ESTER	PFC	ESTER	PFC	ESTER	PFC	ESTER	PFC
L.P.	49	58	9	12	23	17	11	6
R.P.	7	17	5	10	3	10	6	11

2.3 Automatic classification with data mining techniques

Parameter definition : 62 parameters.

Intra-phonemic parameters (40): duration, f0, voicing ratio, formants, intensity.

Inter-phonemic parameters (22): duration, f0, formants, intensity, pauses.

Automatic classification : 25 algorithms using Weka data mining software.

Words	et vs. est				à vs. a			
	ESTER		PFC		ESTER		PFC	
	10Best	Mean	10Best	Mean	10Best	Mean	10Best	Mean
All (62)	77.8	71.3	81.1	76.3	71.4	66.3	66.4	61.6
Formants (30)	65.9	62.3	65.3	62.7	67.7	64.3	61.2	58.5
Prosody (32)	77.7	70.9	81.0	77.3	70.6	65.6	65.9	60.7
Intra- (40)	71.3	65.7	70.4	67.0	68.0	64.0	59.3	57.0
Inter- (22)	74.4	69.2	80.5	77.0	70.1	65.5	65.1	60.1

Results highlight importance of prosodic and inter-phonemic parameters

3 Study in progress and futur works

3.1 Study in progress

Explore links between pronunciation variants and syntactic/semantic classes, syllables and prosody in collaboration with:

CEA LIST : Morpho-syntactic tagging (focus, topic, modality, boundaries)

LIMSI-RITEL : semantic tagging (focus)

3.2 Futur works

Studies at syllabic, prosodic levels.

Chunking and syntactic tagging of audio transcribed corpora.

Automatic learning of acoustic syntactic and prosodic features.

Acoustic syntactic and prosodic model training.

Reference tagged corpora creation for evaluation.

4 Bibliography

[1] R, Nemoto, M, Adda-Decker, I, Vasilescu, *Fouille de données audio pour la discrimination automatique de mots homophones*. In EGC 2008: 445-456. Sophia-Antipolis.

[2] R, Nemoto, I, Vasilescu, M, Adda-Decker, *Speech Errors on Frequently Observed Homophones in French: Perceptual Evaluation vs Automatic Classification*. In Proceedings of the Sixth International LREC'08, ELRA, Marrakech, Morocco.

[3] R, Nemoto, I, Vasilescu, M, Adda-Decker, *Mots fréquents homophones en français: analyse acoustique et classification automatique par fouille de données*. In JEP 2008. Avignon, France.