

UNIVERSITÉ DE RENNES 1

DOCTORAL THESIS

**Characterisation and generation of
expressivity in function of speaking styles
for audiobook synthesis**

Author:

Aghilas SINI

Supervisor:

Dr. Damien LOLIVE

*A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy*

in the

Expression
Enssat

January 10, 2017

“Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism.”

Dave Barry

Université de Rennes 1

Abstract

Enssat
Enssat

Doctor of Philosophy

**Characterisation and generation of expressivity in function of speaking styles for
audiobook synthesis**

by Aghilas SINI

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

Chapter 1

State of the Art

1.1 Expressive Speech Synthesis

In **doukhan:tel-01161731**

- Une des propositions centrales de ce travail est de définir, utiliser et mesurer l'impact de structures linguistiques opérant au delà du niveau de la phrase, par opposition aux approches opérant sur des phrases isolées de leur contexte.
- segmentation des contes en épisodes.
- détection des citations directes.
- actes de dialogue.
- modes de communication.
- Les relations entre les annotations linguistiques et les propriétés prosodiques observées dans le corpus de parole sont décrites et modélisées
- Mis à part les choix lexicaux, la transmission d'images, d'émotions et d'intuitions se fait principalement via un processus de communication non verbale, qui peut se décrire par des intonations, des postures et des mimiques
- générer des instructions prosodiques permettant d'améliorer l'expressivité de la parole synthétisée (livres audio)
- thématique
 - la linguistique de corpus
 - le traitement automatisé du langage
 - le traitement du signal
 - la modélisation prosodique
 - la synthèse de parole.

1.2 Audio Books

Nunc posuere quam at lectus tristique eu ultrices augue venenatis. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aliquam erat volutpat. Vivamus sodales tortor eget quam adipiscing in vulputate ante ullamcorper. Sed eros ante, lacinia et sollicitudin et, aliquam sit amet augue. In hac habitasse platea dictumst.

1.3 SynpaFlex Project

Morbi rutrum odio eget arcu adipiscing sodales. Aenean et purus a est pulvinar pellentesque. Cras in elit neque, quis varius elit. Phasellus fringilla, nibh eu tempus venenatis, dolor elit posuere quam, quis adipiscing urna leo nec orci. Sed nec nulla auctor odio aliquet consequat. Ut nec nulla in ante ullamcorper aliquam at sed dolor. Phasellus fermentum magna in augue gravida cursus. Cras sed pretium lorem. Pellentesque eget ornare odio. Proin accumsan, massa viverra cursus pharetra, ipsum nisi lobortis velit, a malesuada dolor lorem eu neque.

1.3.1 Phonestyle in french

In (AnneCatherine2010PhoneStyle)

1.3.2 Emphatic Stresses

In (Brognaux2014SynthesizingSC)

1.3.3 Sport Commentaries

(brognaux)

1.3.4 roots tool

(chevelu:hal-00974628)

1.3.5 Prosody Modeling Techniques

In (key:article) show the the weakness and advantage of various prosody modeling techniques, also there impact on the speech synthesis systems as well as on emotional speech synthesis systems.

Neural Network based for modeling prosody

In (RamuReddy:2016:PMS:2839529.2839900), the authors show that using PCPA(positional, contextual, phonological and articulatory) as input features of feed forward neural nets it's possible to predict prosody parameters such as Intensity, intonation and duration in the syllable level.

Prosody in Speech Synthesis

In (hirose2015speech), they present different of view, for prosody modeling for getting a High Speech Quality.

In (Rao:2012:PPT:2222511)

1.3.6 data annotation

In (boeffard)

- Définir des informations linguistiques utiles pour améliorer la synthèse
- Concevoir des systèmes permettant d'extraire automatiquement ces informations linguistiques à partir du texte

- Mettre au point des règles de mise en correspondance des informations linguistiques en instructions prosodiques

Chapter 2

Characterization Textual Corpora

2.1 Expressive Speech Labeling

- caractérisation des textes des audios l'aide de structures linguistiques permettant de décrire le texte au delà du niveau de la phrase : structure narrative, référencement des mentions des personnages, citations directes attribuées.

2.2 NEB Corpus

NEB Corpus is free french database "audio books", mono-speaker here the following audio book title:

TABLE 2.1: corpora

carmen	duration	resume of book	prosody
cerfvolant	duration	resume of book	prosody
chantsmaldoror	duration	resume of book	prosody
contes	duration	resume of book	prosody
contescruels	duration	resume of book	prosody
fableslafontaine	duration	resume of book	prosody
filles	duration	resume of book	prosody
germinal	duration	resume of book	prosody
infernaliana	duration	resume of book	prosody
lettresPersanes	duration	resume of book	prosody
miserables	duration	resume of book	prosody
mmeBovary	duration	resume of book	prosody
mystères	duration	resume of book	prosody
poemes	duration	resume of book	prosody
romanmomie	duration	resume of book	prosody
vampire	duration	resume of book	prosody
venus	duration	resume of book	prosody
voyagecentreterre	duration	resume of book	prosody

Dans le tableau ?? page ??.

2.2.1 labeling data

- **.phlab** phonological and phonology labeling.
- **.proslab** prosody labeling.
- **.emolab** emotion labeling.

- **.perslab** personality labeling.