

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

December 14, 2016

Declaration of Authorship

I, Aghilas SINI, declare that this thesis titled, “Characterisation and generation of expressivity in function of speaking styles for audiobook synthesis” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

“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, 2013

- 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 (Simon Anne-Catherine and Jean-Philippe, 2010)

1.3.2 Emphatic Stresses

In (Sandrine, Thomas, and Marco, 2014)

1.3.3 Sport Commentaries

(Brognaux, Picart, and Drugman, 2014)

1.3.4 roots tool

(Chevelu, Lecorvé, and Lolive, 2014)

1.3.5 Prosody Modeling Techniques

In (C and P, 2012) 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 (Ramu Reddy and Sreenivasa Rao, 2016), 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 (Hirose and Tao, 2015), they present different of view, for prosody modeling for getting a High Speech Quality.

In (Rao, 2012)

- 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 2.1 page 3.

2.2.1 labeling data

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

- **.perslab** personality labeling.

Bibliography

- Brognaux, Sandrine, Benjamin Picart, and Thomas Drugman (2014). "Synthèse de commentaires sportifs: intégration d'une annotation prosodique à deux niveaux dans un synthétiseur HMM". In: *Nouveaux Cahiers de la Linguistique Française*. Vol. 31, p. 335. URL: <http://hdl.handle.net/2078.1/150969>.
- C, Rajeswari K and Uma Maheswari P (2012). "Article: Prosody Modeling Techniques for Text-to-Speech Synthesis Systems ? A Survey". In: *International Journal of Computer Applications* 39.16. Full text available, pp. 8–11.
- Chevelu, Jonathan, Gwénolé Lecorvé, and Damien Lolive (2014). "ROOTS: a toolkit for easy, fast and consistent processing of large sequential annotated data collections". In: *Language Resources and Evaluation Conference (LREC)*. Reykjavik, Iceland. URL: <https://hal.inria.fr/hal-00974628>.
- Doukhan, David Doukhan (2013). "Expressive speech synthesis beyond the level of the sentence : the children tale usecase : tale corpora design and analysis for expressive speech synthesis". Theses. Université Paris Sud - Paris XI. URL: <https://tel.archives-ouvertes.fr/tel-01161731>.
- Hirose, Keikichi and Jianhua Tao (2015). *Speech Prosody in Speech Synthesis: Modeling and generation of prosody for high quality and flexible speech synthesis*. Springer.
- Ramu Reddy, V. and K. Sreenivasa Rao (2016). "Prosody Modeling for Syllable Based Text-to-speech Synthesis Using Feedforward Neural Networks". In: *Neurocomput.* 171.C, pp. 1323–1334. ISSN: 0925-2312. DOI: [10.1016/j.neucom.2015.07.053](https://doi.org/10.1016/j.neucom.2015.07.053). URL: <http://dx.doi.org/10.1016/j.neucom.2015.07.053>.
- Rao, K. Sreenivasa (2012). *Predicting Prosody from Text for Text-to-Speech Synthesis*. Springer Publishing Company, Incorporated. ISBN: 1461413370, 9781461413370.
- Sandrine, Brognaux, Drugman Thomas, and Saerens Marco (2014). "Synthesizing Sports Commentaries: One or Several Emphatic Stresses?" In: *Nick Campbell, Dafydd Gibbon & Daniel Hirst (eds.), Social and Linguistic Speech Prosody. Proceedings of the 7th international conference on Speech Prosody*, pp. 270–274. URL: <http://hdl.handle.net/2078.1/150964>.
- Simon Anne-Catherine Auchlin Antoine, Avanzi Mathieu and Goldman Jean-Philippe (2010). "Les phonostyles: une description prosodique des styles de parole en français". In: *Abecassi, M. & G. Ledegen, Les voix des Français. En parlant, en écrivant, Peter Lang*, pp. 71–88. URL: <http://hdl.handle.net/2078.1/83646>.