This book presents the development of a prosody system for text-to-speech (TTS) applications. The prosody is responsible for a communicative intention and guarantees some naturalness in the uttered speech. The prosodic features consist in the imposition of the timing, characterized by the segmental durations and pauses, the intonation, characterized by the fundamental frequency (F0) curve, and by the intensity curve. The proposed prosody model consists of several sub-models, namely, the duration model to predict the segmental durations and the model to predict the F0 pattern. The segmental durations model consists of one ANN carefully selected concerning its architecture and type as well as the set of input features with the objective of minimizing the error between predicted and measured durations. One alternative model, is based on same considerations but uses one dedicated ANN for each phoneme. The alternative model, with dedicated ANNs, improved the final performance. The proposed model to predict the F0 contour is based on the Fujisaki model and consists of two sub-models. One predicts the Phrase Commands parameters and the other predicts the Accent Commands parameters.

João Paulo Teixeira

Prof. João P. Teixeira obtained his PhD in Electrical and Computers Engineering at FEUP in 2004 and worked in several projects related with TTS and Prosody. He teaches at the IPB since 1995 mainly in the area of Signal Processing and is author of several scientific publications in journals and conference proceedings about signal processing and ANN.

978-3-659-16277-0

Prosody Generation Model for TTS Systems

Teixeira

João Paulo Teixeira

Prosody Generation Model for TTS Systems

Segmental Durations and F0 Contours with Fujisaki Model

LAP LAMBERT Academic Publishing

João Paulo Teixeira

Prosody Generation Model for TTS Systems

João Paulo Teixeira

Prosody Generation Model for TTS Systems

Segmental Durations and F0 Contours with Fujisaki Model

LAP LAMBERT Academic Publishing

Impressum / Imprint
Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Bibliographic information published by the Deutsche Nationalbibliothek: The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Coverbild / Cover image: www.ingimage.com

Verlag / Publisher:
LAP LAMBERT Academic Publishing
ist ein Imprint der / is a trademark of
AV Akademikerverlag GmbH & Co. KG
Heinrich-Böcking-Str. 6-8, 66121 Saarbrücken, Deutschland / Germany
Email: info@lap-publishing.com

Herstellung in Deutschland (siehe letzte Seite) /
Printed in the U.S.A. / U.K. by (see last page)
ISBN: 978-3-659-16277-0

Zugl. / Approved by: Bragança, Polytechnic Institute of Bragança, Diss., 2012

Table of Contents

Abbreviations

Aa – Amplitude of AC;
ABU – Acoustic Building Unit;
AC – Accent Command;
ANN – Artificial Neural Network;
Ap – Magnitude of phrase command;
Ca – ANN that predicts the amplitude of the AC;
CA – ANN that predicts the existence of AC associated to the syllable;
EP – European Portuguese;
F0 – Fundamental frequency;
FEUP – Faculty of Engineer of University of Porto;
FEUP-TTS – FEUP Text-To-Speech system;
MOS – Mean Opinion Score;
PC – Phrase Command;
r – Linear correlation coefficient;
rmse – Root mean squared error;
std – Standard deviation;
T0 – Onset time of PC;
T0a – Anticipation of PC;
$T0_E$ – Beginning of accent group where PC was inserted;
T1 – Onset time of AC;
T1a – Anticipation of the onset time of the AC;
T2 – Offset time of AC;
T2a – Anticipation of the offset time of the AC;
TPML – Text Processing Markup Language;
TTS – Text-To-Speech;
δ – Mean absolute error;
σ – Standard deviation.

1 Introduction

This introductory chapter makes a short overview of what is prosody and describes the motivations and objectives for this work. The FEUP-TTS system for European Portuguese, which will be, in first instance, the host of the proposed prosody model, is briefly described. Finally an overview of this document and a reference to the original contributions are made.