SpringerBriefs in Electrical and Computer Engineering

Speech Technology

Series editor

Amy Neustein, Fort Lee, NJ, USA

Editor's Note

The authors of this series have been hand-selected. They comprise some of the most outstanding scientists—drawn from academia and private industry—whose research is marked by its novelty, applicability, and practicality in providing broad based speech solutions. The SpringerBriefs in Speech Technology series provides the latest findings in speech technology gleaned from comprehensive literature reviews and empirical investigations that are performed in both laboratory and *real life* settings. Some of the topics covered in this series include the presentation of real life commercial deployment of spoken dialog systems, contemporary methods of speech parameterization, developments in information security for automated speech, forensic speaker recognition, use of sophisticated speech analytics in call centers, and an exploration of new methods of soft computing for improving human-computer interaction. Those in academia, the private sector, the self service industry, law enforcement, and government intelligence, are among the principal audience for this series, which is designed to serve as an important and essential reference guide for speech developers, system designers, speech engineers, linguists and others. In particular, a major audience of readers will consist of researchers and technical experts in the automated call center industry where speech processing is a key component to the functioning of customer care contact centers.

Amy Neustein, Ph.D., serves as Editor-in-Chief of the International Journal of Speech Technology (Springer). She edited the recently published book "Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics" (Springer 2010), and serves as quest columnist on speech processing for Womensenews. Dr. Neustein is Founder and CEO of Linguistic Technology Systems, a NJ-based think tank for intelligent design of advanced natural language based emotion-detection software to improve human response in monitoring recorded conversations of terror suspects and helpline calls. Dr. Neustein's work appears in the peer review literature and in industry and mass media publications. Her academic books, which cover a range of political, social and legal topics, have been cited in the Chronicles of Higher Education, and have won her a pro Humanitate Literary Award. She serves on the visiting faculty of the National Judicial College and as a plenary speaker at conferences in artificial intelligence and computing. Dr. Neustein is a member of MIR (machine intelligence research) Labs, which does advanced work in computer technology to assist underdeveloped countries in improving their ability to cope with famine, disease/illness, and political and social affliction. She is a founding member of the New York City Speech Processing Consortium, a newly formed group of NY-based companies, publishing houses, and researchers dedicated to advancing speech technology research and development.

More information about this series at http://www.springer.com/series/10043

Moses Effiong Ekpenyong
Editor

Human Language Technologies for Under-Resourced African Languages

Design, Challenges, and Prospects

Springer

Editor
Moses Effiong Ekpenyong
Department of Computer Science
University of Uyo
Uyo, Akwa Ibom State, Nigeria

ISSN 2191-8112 ISSN 2191-8120 (electronic)
SpringerBriefs in Electrical and Computer Engineering
ISSN 2191-737X ISSN 2191-7388 (electronic)
SpringerBriefs in Speech Technology
ISBN 978-3-319-69958-5 ISBN 978-3-319-69960-8 (eBook)
https://doi.org/10.1007/978-3-319-69960-8

Library of Congress Control Number: 2017960943

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

Speech technology has become almost taken for granted in everyday life. Fundamental techniques of automatic speech recognition (ASR), text to speech synthesis (TTS), and speaker recognition and verification have become available as components of commercial interactive search and consumer service agents, which are implemented as smart loudspeakers and supported by complex databases processed with machine learning techniques. Older application fields of public announcements in travel hubs and dictation software and reading applications for the visually impaired are becoming more widespread. Increasingly, speech is understood not only as spoken language but as a multimodal complex of parallel synchronised data streams of audio-visual information from spoken language itself together with body movement: facial mimicry, manual gesture, and posture.

To a large extent, the world of speech technology applications is still focused on the languages of regions with major research and development resources, such as the European Union, North America, India, China, and Japan, but the proceedings of international conference series such as the *Language Resources and Evaluation Conferences* (*LREC*), *InterSpeech*, the *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), the *Language and Technology Conference* (LTC), and *Oriental COCOSDA* have shown a rapid increase in papers on the less resourced languages of Asia, Africa, and the Americas. Even so, these developments are largely driven by technologies developed in the wealthier nations on the basis of their standard, 'commercially interesting' languages.

A central part of the speech system development process is the creation of language and speech resources for the languages concerned. This central part is to a large extent also the most complex part. First, the process splits into many different subprocesses: the pre-recording process of data resource design, the scenario-specific recording processes, and the post-recording procedures of annotation, analysis with machine learning procedures, evaluation, archiving, and dissemination. Second, for each of the subprocesses, the appropriate tool resources need to be developed. A key point to remember is that data type and quality are essentially functions of the tools used in the acquisition and development process, just as

knowledge, in general, is a function of the empirical and formal methods of discovery which are employed.

In the context of speech technologies for the languages of developing nations, there are a number of goals which are currently being addressed in different parts of the global speech technology community. One set of goals may be regarded as a spin-off from the speech technology goals themselves: the use of speech technology tool resources for harvesting data for language documentation, language maintenance by local societies, and indeed in some cases the revival of moribund languages. The fulfilment of these goals meets with many obstacles: traditional scepticism towards new technologies; the desire for prosperity and the conviction that local minority languages do not provide a viable way forward in this direction; problems with providing a reliable infrastructure for the archiving and dissemination of speech technology resources; and lack of financial support for personnel, equipment, and institutional or commercial status for system developers.

The system development goals for the languages of developing nations also differ in complex ways from system development goals for the more highly resourced languages: many languages for which systems have been developed belong to the Indo-European family of languages which have spread through Europe, North and South America, and South Asia, and which differ in many details but also share many features of sound patterns, rhythms and melodies, and word and sentence structure. There are many other types of language in other language families, of which there are many. For the Sino-Tibetan languages, the models developed in the extensive speech technology research being pursued in China are a strong foundation, in the sense that the typology of morphologically isolating tone languages is being dealt with by technologies for Mandarin and Cantonese. But the Niger-Congo languages of Africa have so far only a few isolated speech technology developments, most prominently in North Africa for Arabic and in South Africa for applications to the official languages of the region. In East Africa, developments for Swahili, a non-tonal language, are only partially generalisable to the tonal languages of West and Central Africa. The most active centres for tonal Niger-Congo languages are located in Nigeria, partly in Ibadan, in the African Language Technology Initiative, mainly centred on Yoruba, and in Uyo, mainly centred on Ibibio, and conducted by the authors of this volume.

Overriding issues which need to be overcome in the development of speech technologies for the languages of developing countries can be listed in the *CESAF* criteria. The resources and technologies need to be

- *Comprehensive* with respect to the application domain
- *Effective* in terms of human and computing resources
- *State-of-the-art*, not only intellectually, but also in terms of big data processing, machine learning, and artificial intelligence, not simply the latest internet-dependent software and hardware
- *Affordable* in the sense of being compatible with older computing facilities
- *Fair* with essential involvement of local communities, developers, universities and other research facilities, and companies

The authors of this volume have set themselves the goal of fulfilling the above criteria in order to face the challenge of integrating local languages, in this case in the region of South East Nigeria, into the digital community. The individual chapters address issues in system development and its relation to the digital economy, and to digital services for health and education. These commendable endeavours are already showing fruit and will certainly develop into valuable contributions to overcoming the digital divide. The contributions will provide effective data and tool resource models, not only for Nigeria but also for areas where other Niger-Congo languages are spoken. On this basis, integration of West African speech technology development into the international speech technology community is now well on its way.

Emeritus Professor of English and
General Linguistics at Bielefeld University

Dafydd Gibbon

Preface

Human Language Technology (HLT) also known as Language Technology (LT) is a growing interdisciplinary field that closely connects related sub-disciplines of Linguistics, Psychology, Philosophy, Computer Science, Engineering, Mathematics and Statistics. HLT is naturally induced by Artificial Intelligence (AI) – a term colloquially applied when a machine mimics *cognitive* functions that humans associate with or other aspects of the human mind such as *learning* and *problem solving*.

A SWOT: Strengths-Weaknesses-Opportunities-Threats analysis of HLT indicates that over the years, HLT research has experienced great wave of optimism, with the provision of intelligent tools and methodologies for mining *big data* via the World Wide WEB (WWW) as well as improving at remarkable space, algorithms and applications dedicated to human speech and communication. Further, the quest for comprehensive solution to problems reveals that only an interdisciplinary or agent-based approach with agile methodology is acceptable. HLT also provides far more employment opportunities than are available in traditional academic research, because of its industrial applications. Yet, despite the huge progress in the field, slow developments to extend HLT beyond its frontiers and potentials, subsist – mainly due to failure of related disciplines, industries and stakeholders, to cohesively unite or communicate effectively with each other for a common purpose, thus, posing serious threats to funding justification.

The areas of HLT addressed in this book include speech synthesis, speaker recognition, knowledge representation and spoken language processing. Each chapter addresses an area of HLT. Chapter 1 documents the development of an adaptive synthesis front end for African tone language systems using hidden Markov model (HMM) technique. The template-based front end, though heavily supervised is currently being refined to ensure seamless replicability to a multitude of languages as well as code (re-use) flexibility. In spite of the numerous benefits of speech synthesis, its application is yet to flourish the African domain. Hence, this chapter encourages the development of speech resources to improve the technological status of African tone languages.

Chapter 2 offers an in-depth assessment of speaker variability in speaker recognition systems. It exploits Machine Learning (ML) of relevant acoustic features under degraded/sub-optimal conditions, to demonstrate its feasibility for an under-resourced African tone language. Inspired by the success of voice recognition software such as *Siri* on mobile platforms, companies are itching to place speech interfaces everywhere, and within the next couple of years, voice interfaces will be much more pervasive and powerful. Hence, this chapter provides a useful methodology to assess the suitability of speech features for voice/speaker recognition systems.

Chapter 3 proposes an Ontology Driven application: ODapp, with dynamic framework for spatial context analysis and efficient knowledge representation of multilingual Speech Language Therapy (SLT). SLT is invaluable for the treatment of speech and language disorders, but precise estimate of the number of persons living with such conditions is difficult to obtain, not to talk of doing this across several language domains. This study is therefore apt, as it represents a pioneering initiative within the Nigerian domain. The present research is work in progress and is expected to contribute in improving the poor healthcare services currently experienced in Nigeria, as well as satisfy relevant Sustainable Development Goals (SDGs).

A Spoken-Computer Aided Language Learning (CALL) system is developed in Chap. 4 to demonstrate the application of spoken language processing, a subfield of Human Computer Interaction (HCI). HCI is becoming so faithful at creating interactive products that are easy, flexible and simple for all. However, owing to its multidisciplinary nature and the different value systems of interface users from various backgrounds and experiences, it is highly challenging for designers to create applications which are usable and affordable to such a heterogeneous set of users. An interactive framework driven by speech technology is exploited to enable language learning, as well as protect, preserve and revitalize under-resourced languages.

Finally, this book is designed for research students and staff, language experts, as well as *those behind the curtains* who would wish to explore the growing field of HLTs.

Contents

1 Adaptive Template-Based Front End for Tone Language Speech Synthesis 1
Moses Effiong Ekpenyong

2 Intra-Speaker Variability Assessment for Speaker Recognition in Degraded Conditions: A Case of African Tone Languages 31
Moses Effiong Ekpenyong, Udoinyang G. Inyang, Mercy E. Edoho, and Eno-Abasi E. Urua

3 Towards Ontology-Driven Application for Multilingual Speech Language Therapy 85
Patience U. Usip and Moses Effiong Ekpenyong

4 Ibibio Spoken-CALL System 103
Moses Effiong Ekpenyong, EmemObong O. Udoh, and Nseobong P. Uto

Index 127

About the Author

Moses Effiong Ekpenyong, Ph.D. is a Senior Lecturer in the Department of Computer Science and current Deputy Director of the Centre for Research and Development (CERAD), University of Uyo. He has been involved in research projects and has published widely in his area of specialty: *Speech and Wireless Communications Technology*, with over 100 research publications to his credit. He is a beneficiary of several awards/scholarships and research funding from notable organizations/institutions, such as FGN/Science and Technology Education Post-Basic (STEP-B)-World Bank; Tertiary Education Trust Fund (TETFund), Nigeria; and Outside Echo, UK. He is a reviewer to national and international journals, and belongs to a number of professional bodies, including, Nigeria Computer Society (NCS), Nigerian Mathematical Society (NMS), Institute of Electrical and Electronic Engineers (IEEE), International Speech Communications Association (ISCA), and West African Linguistics Society (WALS).