Foreword

The present state of the art in automatic speech recognition - under certain conditions - permits man-machine-communication by means of spoken language. Provided that speech recognition is tuned to the common native language (target language) of the users, speaker-independent recognition of words from a small vocabulary is feasible, especially if the words are spoken in isolation, and for larger vocabularies at least speaker-dependent recognition performance is satisfactory. The most elaborate up-to-date speech recognition systems manage large vocabularies even in speaker-independent connected speech recognition. However, perfect zero error rates cannot be achieved, and therefore such systems can be used only in applications that to some degree are fault tolerant, e.g. within dialog systems that offer appropriate feedback to the user and allow him to correct recognition errors in a convenient way.

Generally, error rates for speaker-dependent recognition are lower than for speaker-independent recognition, and error rates between these limits are achieved by systems that by default are preset to speaker-independent recognition and can be tailored to a certain speaker by means of a more or less demanding adaptation procedure. For adaptation purposes the prospective user of the system is required to produce a rather large number of certain prescribed utterances. Although some speaker-adaptive systems of this kind are already available, it is still a matter of research, as to how the required number of utterances can be reduced, in order to make the adaptation procedure less demanding for the prospective user. However, at several places more ambitious research work is directed towards speech recognition systems that continuously adapt to the current speaker without requiring a separate adaptation phase at each change between speakers. The work that is presented here is a major step towards such a system and - due to a remarkable new approach - adaptation can be successful even in the case of a non-native speaker with a foreign accent.

In an elaborate speech recognition system there are several knowledge sources among which the pronunciation dictionary is of special interest for measures against non-native accents. Usually, this dictionary denotes the native pronunciations for all items of the vocabulary. In oder to make the system ready for non-native accents, also the respective non-native pronunciations must be entered into the pronunciation dictionary. Up to now these additional entries either had to be specified by an expert for the special pair of target language and foreign language or had to be extracted automatically from a large number of spoken examples. The new approach is based upon the idea of processing the vocabulary and pronunciation dictionary of the target language with special attention to the phoneme inventory of the foreign language. Thus, any desired pair of target language and foreign language can be conveniently managed without the help of special experts and without any spoken examples. This is the most important among the contributions of this work to the fundamentals for the development of future speech recognition systems.

November 2002 E. Paulus

Preface

Speech recognition technology is being increasingly employed in human-machine interfaces. Two of the key problems affecting such technology, however, are its robustness across different speakers and robustness to non-native accents, both of which still create considerable difficulties for current systems.

In this book methods to overcome these problems are described. A speaker adaptation algorithm that is based on Maximum Likelihood Linear Regression (MLLR) and that is capable of adapting the acoustic models to the current speaker with just a few words of speaker specific data is developed and combined with confidence measures that focus on phone durations as well as on acoustic features to yield a semi-supervised adaptation approach. Furthermore, a specific pronunciation modelling technique that allows the automatic derivation of non-native pronunciations without using non-native data is described and combined with the confidence measures and speaker adaptation techniques to produce a robust adaptation to non-native accents in an automatic speech recognition system.

The aim of this book is to present the state of the art in speaker adaptation, confidence measures and pronunciation modelling, as well as to show how these techniques have been improved and integrated to yield a system that is robust to varying speakers and non-native accents.

The faculty of electrical engineering of the Technical University Carolo-Wilhelmina of Braunschweig has accepted this book as a dissertation and at this point I would like to take the opportunity to thank all those who supported me during this research work. First of all I would like to thank Prof. Dr.-Ing. Erwin Paulus for his valuable suggestions and his support. Likewise I would like to thank Prof. Dr.-Ing. Günther Ruske from the Technical University of Munich for giving the second opinion.

The research presented in this book was conducted while I was working at Sony in Stuttgart and I would like to express my gratitude that I got the permission to conduct and publish this research work.

I am also indebted to my colleagues, who supported me a lot by being very cooperative and helpful. I would like especially to thank Dr. Krzysztof Marasek and Andreas Haag for their co-operation in the field of confidence measures, Dr. Stefan Rapp for the valuable discussions and Dipl. Ling. Manya

Sahakyan for conducting the accompanying experiments in the area of pronunciation modelling.

Furthermore I would like to thank Dr. Elmar Noeth, Dr. Richard Stirling-Gallacher, Dr. Franck Giron and particularly Dr. Roland Kuhn for their many very helpful comments, which played an important part in improving this book.

Special thanks go to Dr. Ralf Kompe, whose support was an important contribution to this book and who, in spite of his many other duties, always found time to support me.

Most especially, I would like to thank my dear parents, Waldemar and Renate Goronzy, without whose affectionate upbringing my entire education and this book would not have been possible and who provided me with a set of values I could not have gained through any other academic education process.

Fellbach, September 2001 Silke Goronzy