UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

SPARSE CODING FOR SPEECH RECOGNITION

By

Willem Jacobus Smit

Submitted in partial fulfilment of the requirements for the degree

Philosophiae Doctor (Electronic Engineering)

in the

Faculty of Engineering, the Built Environment and Information Technology

at the

UNIVERSITY OF PRETORIA

Promoter: Professor E. Barnard

April 2008

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ACKNOWLEDGEMENTS

I enjoyed the work towards this thesis very much. For this I would like to thank several people. Firstly, thanks to Etienne Barnard, he is an excellent mentor and study leader. I would also like to thank Johann Holm and Liesbeth Botha for their guidance in the first few years I was working on my PhD. Also thanks to my wife Janine, she is very supportive and understanding. I look forward to taking on similar tasks with her support.

I would like to thank the National Research Foundation and the University of Pretoria for their financial support.

Lastly, praise to the Lord for creating a wonderful universe that we can explore.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

SPARSE CODING FOR SPEECH RECOGNITION

BY WILLIE SMIT

PROMOTER: PROFESSOR E. BARNARD

DEPARTMENT OF ELECTRICAL, ELECTRONIC AND COMPUTER ENGINEERING

PHILOSOPHIAE DOCTORAL

The brain is a complex organ that is computationally strong. Recent research in the field of neurobiology help scientists to better understand the working of the brain, especially how the brain represents or codes external signals. The research shows that the neural code is sparse. A sparse code is a code in which few neurons participate in the representation of a signal.

Neurons communicate with each other by sending pulses or spikes at certain times. The spikes send between several neurons over time is called a spike train. A spike train contains all the important information about the signal that it codes. This thesis shows how sparse coding can be used to do speech recognition. The recognition process consists of three parts. First the speech signal is transformed into a spectrogram. Thereafter a sparse code to represent the spectrogram is found. The spectrogram serves as the input to a linear generative model. The output of the model is a sparse code that can be interpreted as a spike train. Lastly a spike train model recognises the words that are encoded in the spike train.

The algorithms that search for sparse codes to represent signals require many computations. We therefore propose an algorithm that is more efficient than current algorithms. The algorithm makes it possible to find sparse codes in reasonable time if the spectrogram is fairly coarse.

The system achieves a word error rate of 19% with a coarse spectrogram, while a system based on Hidden Markov Models achieves a word error rate of 15% on the same spectrograms.

Key terms: sparse code, sparse code measurement, speech recognition, linear generative model, spike train, spike train classification, mathematical optimization, spectrogram, overcomplete dictionary, dictionary training.

YLKODES VIR SPRAAKHERKENNING

DEUR WILLIE SMIT

PROMOTOR: PROFESSOR E. BARNARD

DEPARTMENT VAN ELEKTRIESE, ELEKTRONIESE EN REKENAARINGENIEURSWESE

PHILOSOPHIAE DOCTORAL

Die brein is ’n komplekse orgaan wat berekeningsgewys baie sterk is. Onlangse navorsing in die veld van neurobiologie help wetenskaplikes om die werking van die brein beter te verstaan, veral hoe die brein eksterne seine voorstel of kodeer. Die navorsing wys onder andere dat die neurale kode yl is. ’n Ylkode is ’n kode waar min breinselle deelneem om ’n spesifieke sein voor te stel.

’n Breinsel kommunikeer met ander selle deur pulse op spesifieke tye te stuur. Die pulse tussen verskeie breinselle oor tyd word ’n pulsreeks genoem. Die pulsreeks bevat al die belangrikke inligting oor die sein wat dit kodeer. Hierdie tesis wys hoe ylkodes gebruik kan word om spraakherkenning te doen. Die herkenningsproses bestaan uit drie dele. Eers word ’n spraaksein voorgestel as ’n spektrogram. Daarna word ’n ylkode gesoek wat weer die spektrogram voorstel. Die spektrogram dien as inset tot ’n lineêr–genererende model. Die model se uitset is ’n ylkode wat geinterpreteer kan word as ’n pulsreeks. Laastens herken ’n pulsreeksmodel die woorde wat voorgestel word in die pulsreeks.

Die algoritmes wat ylkodes soek om seine voor te stel verg baie berekeninge. Ons stel daarom ’n algoritme voor wat meer effektief is as huidige algoritmes. Die algoritme maak dit moontlik om ylkodes te vind binne ’n redelike tyd as die spektrogram taamlik grof is.

Met ’n growwe spektrogram behaal die herkenningsstelsel ’n woordfouttempo van 19%, terwyl ’n stelsel gebaseer op verskuilde Markovmodelle (“Hidden Markov Models”) ’n woordfouttempo van 15% behaal vir dieselfde voorstelling.

Sleutelterme: ylkode, ylkode maatstaf, spraakherkenning, lineêr–genererende model, pulsreekse, pulsreeksherkenning, wiskundige optimering, spektrogram, oorvolledige woordeboek, woordeboek afrigting.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

TABLE OF CONTENTS