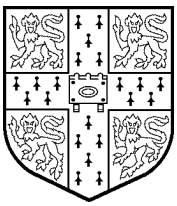
Speech Processing with Linear and Neural Network Models

Tina-Louise Burrows



Cambridge University Engineering Department
Trumpington Street
Cambridge CB2 1PZ
England

This dissertation is submitted for consideration for the degree of Doctor of Philosophy at the University of Cambridge

Summary

This dissertation investigates some aspects of speech processing using linear models and single hidden layer neural networks. The study is divided into two parts which focus on speech modelling and speech classification respectively.

The first part of the dissertation examines linear and nonlinear vocal tract models for synthesising high quality speech with adjustable pitch. A source-filter framework for analysis and synthesis is used, in which the source is a representation of the glottal volume velocity waveform. Two families of linear model are considered, ARX (autoregressive with external input) and OE (output error). Their performance in estimating vocal tract transfer functions is compared on synthetic speech data, and the difference is explained in terms of the parameter estimation procedure, the frequency distribution of bias in the estimate and the assumptions about the spectrum of the noise in the vocal tract system. The noise spectrum for ARX models is shown to be perceptually significant for speech synthesis applications because it exploits auditory masking. Methods for improving poor quality syntheses from OE models are proposed. Nonlinear vocal tract models, implemented as feed-forward or recurrent neural networks, are investigated. Methods for initialising networks from linear models are developed. A modified recurrent architecture is introduced which permits initialisation from ARX models. The use of regularization, for imposing continuity between models of adjacent speech segments, and learning rate adaptation, for improving back-propagation training, are discussed. For synthesising real speech utterances, an audio tape demonstrates that ARX models produce the highest quality synthetic speech and that the quality is maintained when pitch modifications are applied.

The second part of the dissertation studies the operation of recurrent neural networks in classifying patterns of correlated feature vectors. Such patterns are typical of speech classification tasks. The operation of a hidden node with a recurrent connection is explained in terms of a decision boundary which changes position in feature space. The feedback is shown to delay switching from one class to another and to smooth output decisions for sequences of feature vectors from the same class. For networks trained with constant class targets, a sequence of feature vectors from the same class tends to drive the operation of hidden nodes into saturation. It is demonstrated that saturation defines limits on the position of the decision boundary resulting in context-sensitive and context-insensitive regions of the feature space. While saturation persists, it is shown that networks have reduced sensitivity to the order of presentation of feature vectors because movement of the decision boundary is inhibited. To improve this within-class sensitivity, training with ramp-like class targets is investigated. The operation of small recurrent networks is demonstrated for two tasks; classification of speech utterances into voiced and unvoiced segments, and classification of clockwise and anti-clockwise trajectories of vectors produced by two autoregressive processes.

Acknowledgements

I would like to thank everyone in the Fallside Lab for making my time in Cambridge an experience. In particular, I would like to mention Julian for his practical advice, Rob for all his help with the fiddly tape-recording, and Xtof for his patience with my faltering Spanish. Special thanks to my supervisor, Dr. Mahesan Niranjan, for his guidance, and to Dr. Ljung and Dr. Maciejowski for helpful discussions on system identification theory. The biggest thank-you of all goes to my sister, Tanya, for all her love and support, especially while writing up.

This work has been funded by the Science and Engineering Research Council with some useful top-ups from the Engineering Department and Queens' College.

Dedication

To Mum and Dad. Thank you for supporting me in all the mad things I do.

Declaration

This 54,000 word dissertation is entirely the result of my own work and includes nothing which is the outcome of work done in collaboration.

Tina-Louise Burrows
Queens' College
March 20, 1996

Contents

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.