Waveform Interpolation Speech Coder at 4 kb/s

Eddie L. T. Choy

Department of Electrical and Computer Engineering
McGill University
Montréal, Canada

August 1998

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Engineering.

Abstract

Speech coding at bit rates near 4 kbps is expected to be widely deployed in applications such as visual telephony, mobile and personal communications. This research focuses on developing a speech coder based on the waveform interpolation (WI) scheme, with an attempt to deliver near toll-quality speech at rates around 4 kbps. A WI coder has been simulated in floating-point using the C programming language. The high performance of the WI model has been confirmed by subjective listening tests in which the unquantized coder outperforms the 32 kbps G.726 standard (ADPCM) 98% of the time under clean input speech conditions; the reconstructed speech is perceived to be essentially indistinguishable from the original. When fully quantized, the speech quality of the WI coder at 4.25 kbps has been judged to be equivalent to or better than that of G.729 (the ITU-T toll-quality 8 kbps standard) for 45% of the test sentences. Further refinements of the quantization techniques are warranted to bring the coder closer to the toll-quality benchmark. Yet, the existing implementation has produced good quality coded speech with a high degree of intelligibility and naturalness when compared to the conventional coding schemes operating in the neighbourhood of 4 kbps.

Sommaire

Dans un futur proche, le codage de la parole à des taux autour de 4 kbps devrait être largement utilisé dans des applications comme, la teléphonie visuelle, et les communications personnelles et mobiles. Cette recherche a pour but de développer un codeur de parole basé sur l'interpolation d'un signal (abrégé WI pour *waveform interpolation*), avec comme objectif une reconstruction fidèle de la parole à des débits aussi faibles que 4 kbps. Un codeur basé sur le modèle WI a été simulé en arithmétique flottante en utilisant le language C. Les hautes performances du modèle ont été confirmées par des tests d'écoute dans lesquels la qualité de parole du codeur sans quantification est meilleure que le standard 32 kbps G.726 (ADPCM) dans 98% des cas lorsque la parole utilisée au départ était sans bruit. On peut conclure que la synthèse est perçue comme étant essentiellement indifférentialle de la parole originale. Quand les paramètres du codeur sont complètement quantifiés, la qualité de parole du codeur WI à 4.25 kbps a été jugée comme étant équivalente ou meilleure que le G.729 (le standard ITU-T toll-quality 8 kbps) pour 45% des sequences de test. Des améliorations plus poussées des techniques de quantification sont nécessaires pour que le codeur permette une reconstruction encore plus proche de la reconstruction fidèle. Néanmoins, le programme existant a donné de la parole codée de bonne qualité avec un haut degré d'intelligibilité et de naturel comparé aux autres codeurs conventionnels fonctionnant autour de 4 kbps.

Acknowledgments

I would like to express my sincere thanks to my supervisor, Professor Peter Kabal, for his guidance and support throughout my graduate studies at McGill University. Also, I am thankful to Dr. Jacek Stachurski for co-implementing the waveform interpolation speech coder. This research could not have been possible without their technical expertise, critical insight and enlightening suggestions.

Moreover, I acknowledged all my fellow graduate students in the Telecommunications and Signal Processing Laboratory for their encouragement and companionship. Special thanks go to Hossein, Nadim and Khaled who constantly gave me both technical and non-technical advice. I am also obliged to Florence who helped me with the French abstract.

I am thankful to Jianming, Michael, Johnny and Mohammad who participated in the listening tests for this research. The postgraduate scholarship awarded by the Natural Sciences and Engineering Research Council of Canada is appreciated.

My deepest gratitude goes to my fiancée Jane for her love and understanding, and also to our respective families for their continuous support and encouragement in the past two years.

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.