Signals and Communication Technology

More information about this series at http://www.springer.com/series/4748

Tom Bäckström

Speech Coding

with Code-Excited Linear Prediction

Springer

Tom Bäckström
International Audio Laboratories Erlangen
(AudioLabs)
Friedrich-Alexander University
Erlangen-Nürnberg (FAU)
Erlangen
Germany

ISSN 1860-4862 ISSN 1860-4870 (electronic)
Signals and Communication Technology
ISBN 978-3-319-50202-1 ISBN 978-3-319-50204-5 (eBook)
DOI 10.1007/978-3-319-50204-5

Library of Congress Control Number: 2017930277

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

Preface

Depending on your source of reference, there are between 7 and 8 billion mobile phones in use at the time of writing this work. It is a breathtaking technological success. In less than 20 years, mobile telephony has reached all parts of the world and there is an almost universal access to mobile phones for all people in the world. That is not to say that all 7 billion people of the world would have their own phone since many have multiple phones, but a large majority of people does have *access* to a mobile phone. Almost all people in the world have the option of making a (mobile) phone call.

The consequences of this technological leap are as astonishing as the speed with which this leap has happened. In the industrial world, practically every adult has a mobile phone. It is a defining characteristic of our age—to be always reachable, to be always online. It has fundamentally changed our attitude towards communication. Being within reach is no more a thing that needs to be planned, gone are the days when we would agree the time and place when we would talk the next time. We are simply always there.

Improved telecommunication solutions have the potential of increasing the amount of communication between people. This is almost only a good thing. Increased communication improves understanding between people, it increases awareness of society and your surroundings, and has thus an impact on the level of democracy.

In the industrial world, democracy can of course be *nice*, but where mobile telephony has had a much larger impact on democracy is the developing world. It is fairly common for small, poor villages and communities in Africa to share one mobile phone for the whole village. It has proved to be more expensive to build and maintain land-lines than mobile networks, and thus hardly any new land-lines are being built. There are plenty of villages that never had a phone, never had an instantaneous communication channel, before the mobile phone. Where before, people had to walk to the next village to get news from their relatives, they can now simply make a phone call. They can even talk to their relatives who have moved abroad, perhaps even overseas, and get the latest news. Imagine the societal and human impact of that: For the first time in human history, any one person in the

world can connect with anybody else in an instant, independent of geographic location!

As with every success, there is a drawback; the huge steps forward have proved to be the biggest enemy of the technology. For example, it would be easy to make the conclusion that since 7 billion people have access to phones, *it works already* and there is thus not much to be improved. Alas, how often have I not heard this claim, even from scientists in the field!

Fortunately, the same argument can be turned upside down to demonstrate the importance of further improvements to mobile telephony. Imagine the impact that the tiniest of improvements would have, when multiplied by 7 billion users! The largest fixed costs of mobile networks surely lie in the network, where efficiency and costs are directly proportional. A 1% improvement in coding efficiency, in the actual amount of data sent, does not have a large impact on the individual phone, but worldwide, already the reduced energy consumption due to a 1% improvement in efficiency would have a huge impact.

At the same time, as the number of devices capable of speech coding is exponentially increasing, the amount of energy required is also increasing rapidly despite improvements in energy efficiency of CPUs. To make it feasible to use speech codecs on such a large number of devices, it will be necessary to reduce the complexity of speech coding algorithms. This is especially evident with the emerging concept of the internet-of-things, where it is envisioned that all electronic devices would be connected and capable of communication. It is obvious that we need to limit the complexity of codecs to reduce power consumption if they are to be implemented on such a large number of devices.

Moreover, even if only a fraction of the improvement yields lower costs for the end-users, it is not difficult to see its impact on that small African village. 50 cents saved per month is good money.

The other obstacles for mobile telephony imposed by its own success are the economic incentives within standardisation processes. Regrettably, however, the current author is not authorised to pursue an academic discussion of the economic incentives of standardisation any further.

It is then not surprising that during the last decade academic research in speech coding has been scarce. It seems that most of the cutting-edge research happens behind closed, corporate doors and little if any information is leaked out. All that is visible to the outside are the (sometimes) brilliant engineering art-works of speech coding standards, which include hardly any new science. Speech codecs have become very finely tuned machineries which are so complex, that the tiniest modification in any part inevitably breaks something somewhere else.

The purpose of this book is to take a step towards science. The objective is to formalise the most common speech coding tools into a scientific framework, such that their strengths and weaknesses can be assessed in isolation, without the complex interconnections of a full-scale codec. The hope is that this work will give a stable scientific framework which allows for new innovations and paves the way for new break-through technologies.

The emphasis in this book is in understanding *why* and *how* the commonly used methods work. Specific details about standards have for the most parts been intentionally left out to keep the overall structure clear, but ample references are provided for the experts.

The book was originally created as a compendium for the course "Speech Coding" which I taught at the Friedrich-Alexander University Erlangen-Nürnberg (FAU). It is therefore designed to be useful as learning material for students working on a master's degree in speech and audio processing, as well as information or communications technology. In parallel, the book is also meant to provide in-depth information for engineers and scientists working in the field and can thus be used as a handbook on speech coding.

Erlangen, Germany Tom Bäckström
October 2016

Acknowledgements

What I know about practical methods for speech coding I learned in most parts during my time at the International Audio Laboratories Erlangen[1] from 2008 until 2016. Here, I had the privilege of being closely involved in the standardisation of both the MPEG Unified Speech and Audio Codec (standardised 2012) as well as the 3GPP Enhanced Voice Services speech codec (standardised 2014). This environment provided me demanding development tasks in speech coding but simultaneously sufficient time to learn and develop theory as well. With this book I wish to show my gratefulness for my colleagues at the AudioLabs, FAU as well as Fraunhofer.

I have had the pleasure of working on this book with four co-authors, Sascha Disch, Jerémié Lecomte, Christian Uhle and Guillaume Fuchs, each submitting a guest-authored chapter. All of them are the best experts in their respective fields, complementing those areas where I have less experience. I am grateful for their contributions and I think it is obvious that their participation made the book much better and a more balanced whole.

In writing this book, I have received helpful comments and reviews from Tim Fingscheidt, Christian Fischer Pedersen, Goran Markovic, as well as many students of my courses and other colleagues at the AudioLabs. I remain indebted to you all.

Disclaimer: A chapter of this book was removed from the published version due to a claim that it infringes a non-disclosure agreement.

Erlangen, Germany Tom Bäckström
October 2016

[1]International Audio Laboratories Erlangen is a joint institute between Fraunhofer Institute of Integrated Circuits (IIS) and Friedrich-Alexander University Erlangen-Nürnberg (FAU).

Contents

About the Authors

Tom BäckstrÖm, D.Sc. (tech), is now professor at Aalto University, in Helsinki, Finland, but during the preparation of this book he was professor at International Audio Laboratories Erlangen, which is a co-operation unit between the Friedrich-Alexander University Erlangen-Nürnberg (FAU) and Fraunhofer Institute of Integrated Circuits (IIS), both in Erlangen, Germany.

Sascha Disch, Dr.-Ing., is a senior scientist at International Audio Laboratories Erlangen, Fraunofer IIS.

Jérémie Lecomte was a reseacher at Fraunofer IIS during the perparation of this book, but is now at Amazon.

Christian Uhle, Dr.-Ing., is a senior scientist at International Audio Laboratories Erlangen, Fraunofer IIS.

Guillaume Fuchs, Ph.D., is a senior scientist at International Audio Laboratories Erlangen, Fraunofer IIS.

Acronyms

Throughout this work, acronyms and abbreviations have been avoided whenever possible. The motivation is that since many abbreviations are known only within a very specific professional field, using them makes it difficult for the uninitiated to join or even follow a discussion. Since acronyms are, however, still used frequently among engineers and scientists, the following list of acronyms is provided as a reference.

3GPP	3rd generation partnership project (standardisation organisation)
AbS	Analysis by synthesis (optimisation method)
ACELP	Algebraic code-excited linear prediction (speech codec)
ACR	Absolute category rating (subjective evaluation method)
ADPCM	Adaptive pulse code modulation (quantisation method with source model)
AMR	Adaptive multirate (speech coding standard)
AMR-NB	Adaptive multirate narrowband (speech coding standard)
AMR-WB	Adaptive multirate wideband (speech coding standard)
AMR-WB+	Adaptive multirate wideband Plus (speech coding standard)
ARQ	Automatic retransmission request (method for concealment of lost packets)
BBWE	Blind bandwidth extension (advanced speech and audio coding tool)
BPF	Bass postfilter (advanced speech and audio coding tool)
BWE	Bandwidth extension (advanced speech and audio coding tool)
CBR	Constant bitrate (transmission rate)
CCR	Comparison category rating (subjective evaluation method)
CELP	Code-excited linear prediction (speech codec)
CNG	Comfort noise generator (advanced speech coding tool)
DCR	Degradation category rating (subjective evaluation method)
DPCM	Differential pulse code modulation (quantisation method which assumes that signal has low-pass character)

DTX	Discontinuous transmission (coding mode)
ERB	Equivalent rectangular bandwidth (frequency scale corresponding to perceptual accuracy)
EVS	3GPP Enhanced voice services (speech coding standard)
FEC	Forward error correction (method for concealment of lost packets)
FIR	Finite impulse response (filter)
G.718	Frame error robust narrowband and wideband embedded variable bitrate coding of speech and audio from 8-32 kbit/s (speech coding standard)
HNR	Harmonics to noise ratio (measure of signal characteristic)
IID	Independent and identically distributed (random variable)
IIR	Infinite impulse response (filter)
IPA	International phonetic alphabet (phonetic notation system)
ISF	Immittance spectrum frequency (representation)
ISP	Immittance spectrum pair (polynomials)
ITU	International telecommunication union (standardisation organisation)
ITU-D	ITU's Development sector (standardisation organisation)
ITU-R	ITU's Radio communication sector (standardisation organisation)
ITU-T	ITU's Telecommunication standardisation sector (standardisation organisation)
LPC	Linear predictive coding (method)
LSD	Log-spectral distortion (distance measure)
LSF	Line spectrum frequency (representation)
LSP	Line spectrum pair (polynomials)
LTP	Long-term prediction (method)
MDCT	Modified discrete cosine transform (lapped time-frequency transform)
MPEG	Moving picture experts group (standardisation organisation)
MUSHRA	Multiple stimuli with hidden reference and anchor (subjective quality evaluation method)
PCM	Pulse code modulation (quantisation method)
PEAQ	Method for objective measurements of perceived audio quality (objective quality evaluation method)
PESQ	Perceptual evaluation of speech quality (objective quality evaluation method)
PLC	Packet loss concealment (advanced speech coding tool)
POLQA	Perceptual objective listening quality assessment (objective quality evaluation method)
RCELP	Relaxed code-excited linear prediction (low-bitrate speech codec)
SAMPA	Speech assessment methods phonetic alphabet (phonetic notation system)
SBR	Spectral band replication (method for bandwidth extension)

USAC	MPEG Unified speech and audio codec (speech coding standard)
VAD	Voice activity detection (advanced speech coding tool)
VBR	Variable bitrate (transmission rate)
ZIR	Zero input response (part of the filter windowing method)