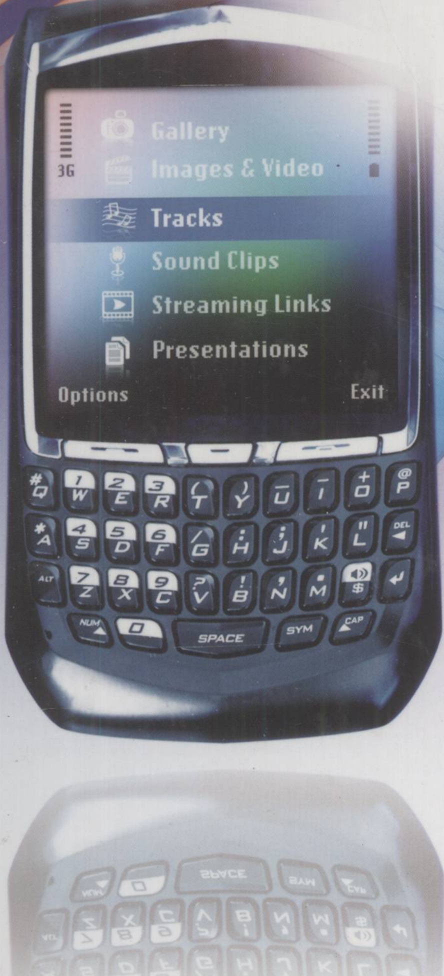


Voice and Audio Compression for Wireless Communications

Second Edition

Lajos Hanzo, F. Clare Somerville and Jason Woodard

WILEY

Voice and Audio Compression for Wireless Communications

Second Edition

Lajos Hanzo
University of Southampton, UK

F. Clare Somerville
picoChip Designs Ltd, UK

Jason Woodard
CSR plc, UK

IEEE
IEEE PRESS
IEEE Communications Society, Sponsor

BICENTENNIAL 1807 WILEY 2007 BICENTENNIAL

John Wiley & Sons, Ltd

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

IEEE Communications Society, Sponsor
COMMS-S Liaison to IEEE Press, Mostafa Hashem Sherif

Library of Congress Cataloging-in-Publication Data

Hanzo, Lajos, 1952-
Voice and Audio Compression for Wireless Communications / L. Hanzo,
F.C.A. Somerville and J.P. Woodard – 2nd ed.
p. cm.
Rev. ed. of: Voice and Audio Compression for Wireless Communications. c2001
Includes bibliographical references and index.
ISBN 978-0-470-51581-5 (cloth : alk. paper)
1. Compressed speech. 2. Speech processing systems. 3. Telecommunication systems.
I. Somerville, F. Clare A. II. Woodard, Jason P. III. Hanzo, Lajos,
1952- Voice compression and communications. IV. Title.
TK7882.S65H35 2007
621.384–dc22

2007011025

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 978-0-470- 51581-5 (HB)

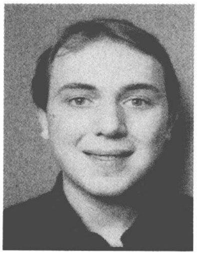
Typeset by the authors using LaTeX software.
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, England.
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Voice and Audio Compression for Wireless Communications

About the Authors

Lajos Hanzo FREng, FIEEE, FIET, DSc received his degree in electronics in 1976 and his doctorate in 1983. During his 30 year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the School of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has co-authored 14 books on mobile radio communications totalling in excess of 10 000 pages, published about 700 research papers, acted as TPC Chair of IEEE conferences, presented keynote lectures and been awarded a number of distinctions. Currently he is directing an academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European IST Programme and the Mobile Virtual Centre of Excellence (VCE), UK. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also an IEEE Distinguished Lecturer of both the Communications Society and the Vehicular Technology Society (VTS). Since 2005 he has been a Governor of the VTS. For further information on research in progress and associated publications please refer to http://www-mobile.ecs.soton.ac.uk.

Clare Somerville (nee Brooks) received the M.Eng in Information Engineering, in 1995, from the University of Southampton, UK. From 1995 to 1998 she performed research into low-bitrate speech coders for wireless communications leading to a PhD in 1999, also from the University of Southampton. From 1998 to 2001 she was with the Global Wireless Systems Research department, Bell Laboratories, Swindon, UK where she undertook research into real-time services over GPRS networks. Since 2001 she has been a Principal Systems Engineer at picoChip Designs Ltd, Bath UK. working on protocol layer aspects in both UMTS and WiMAX wireless systems. Her current interests lie within the picoChip WiMAX product range where she is lead architect for the MAC. She is a member of the 802.16 standards forum and is a registered mentor for Women in SET (Science, Engineering and Technology).



Jason Woodard was born in Northern Ireland in 1969. He received a BA degree in Physics from Oxford University in 1991, and an MSc with distinction in Electronics from the University of Southampton in 1992. In 1995 he completed a PhD in speech coding, also at the University of Southampton, and then held a three year postdoctoral fellowship, researching turbo-coding techniques for the FIRST project within the European ACTS programme. In 1998 he joined the PA Consulting Group in Cambridge, UK and in 1999 he was a founding member of UbiNetics, a supplier of mobile communications test and IP solutions. Currently he is working on the research and development of advanced wireless technologies for CSR, a leading supplier of single-chip wireless devices. Dr Woodard has published widely in wireless communications, including co-authoring two books.

Other Wiley and IEEE Press Books on Related Topics[1]

- R. Steele, L. Hanzo (Ed): *Mobile Radio Communications: Second and Third Generation Cellular and WATM Systems*, John Wiley & Sons, Ltd and IEEE Press, 2nd edition, 1999, ISBN 07 273-1406-8, 1064 pages
- L. Hanzo, F.C.A. Somerville, J.P. Woodard: *Voice Compression and Communications: Principles and Applications for Fixed and Wireless Channels*, IEEE Press and John Wiley & Sons, Ltd, 2001, 642 pages
- L. Hanzo, P. Cherriman, J. Streit: *Wireless Video Communications: Second to Third Generation and Beyond*, IEEE Press and John Wiley & Sons, Ltd, 2001, 1093 pages
- L. Hanzo, T.H. Liew, B.L. Yeap: *Turbo Coding, Turbo Equalisation and Space–time Coding*, John Wiley & Sons, Ltd and IEEE Press, 2002, 751 pages
- J.S. Blogh, L. Hanzo: *Third-Generation Systems and Intelligent Wireless Networking: Smart Antennas and Adaptive Modulation*, John Wiley & Sons, Ltd and IEEE Press, 2002, 408 pages
- L. Hanzo, C.H. Wong, M.S. Yee: *Adaptive Wireless Transceivers: Turbo-Coded, Turbo-Equalised and Space–time Coded TDMA, CDMA and OFDM Systems*, John Wiley & Sons, Ltd and IEEE Press, 2002, 737 pages
- L. Hanzo, L.-L. Yang, E.-L. Kuan, K. Yen: *Single- and Multi-Carrier CDMA: Multi-User Detection, Space–time Spreading, Synchronisation, Networking and Standards*, John Wiley & Sons, Ltd and IEEE Press, June 2003, 1060 pages
- L. Hanzo, M. Münster, T. Keller, B.-J. Choi: *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs and Broadcasting*, John-Wiley & Sons, Ltd and IEEE Press, 2003, 978 pages
- L. Hanzo, S.-X. Ng, T. Keller and W.T. Webb: *Quadrature Amplitude Modulation: From Basics to Adaptive Trellis-Coded, Turbo-Equalised and Space–time Coded OFDM, CDMA and MC-CDMA Systems*, John Wiley & Sons, Ltd and IEEE Press, 2004, 1105 pages

[1]For detailed contents and sample chapters please refer to http://www-mobile.ecs.soton.ac.uk.

- L. Hanzo, T. Keller: *An OFDM and MC-CDMA Primer*, John Wiley & Sons, Ltd and IEEE Press, 2006, 430 pages
- L. Hanzo, F.C.A. Somerville, J.P. Woodard: *Voice and Audio Compression for Wireless Communications*, John Wiley & Sons, Ltd and IEEE Press, 2007, 858 pages
- L. Hanzo, P.J. Cherriman, J. Streit: *Video Compression and Communications: H.261, H.263, H.264, MPEG4 and HSDPA-Style Adaptive Turbo-Transceivers*, John Wiley & Sons, Ltd and IEEE Press, 2007, 680 pages
- L. Hanzo, J.S. Blogh, S. Ni: 3G Systems and HSDPA-Style FDD Versus TDD Networking: Smart Antennas and Adaptive Modulation, John Wiley & Sons, Ltd and IEEE Press, 2007

Preface and Motivation

The Speech Coding Scene

Despite the emergence of sophisticated high-rate multimedia services, voice communications remain the predominant means of human communications, although the compressed voice signals may be delivered via the Internet. The large-scale, pervasive introduction of wireless Internet services is likely to promote the unified transmission of both voice and data signals using the Voice over Internet Protocol (VoIP) even in the third-generation (3G) wireless systems, despite wasting much of the valuable frequency resources for the transmission of packet headers. Even when the predicted surge of wireless data and Internet services becomes a reality, voice remains the most natural means of human communications, although this may be delivered via the Internet.

This book is dedicated to audio and voice compression issues, although the aspects of error resilience, coding delay, implementational complexity and bitrate are also at the centre of our discussions, characterising many different speech codecs incorporated in source-sensitivity matched wireless transceivers. A unique feature of this book is that it also provides cutting-edge turbo-transceiver-aided research-oriented design examples and a chapter on the VoIP protocol.

Here we attempt a rudimentary comparison of some of the codec schemes treated in the book in terms of their speech quality and bitrate, in order to provide a road map for the reader with reference to Cox's work [1,2]. The formally evaluated mean opinion score (MOS) values of the various codecs portrayed in this book are shown in Figure 1.

Observe in the figure that over the years a range of speech codecs have emerged, which attained the quality of the 64 kbps G.711 pulse-code modulation (PCM) speech codec, although at the cost of significantly increased coding delay and implementational complexity. The 8 kbps G.729 codec is the most recent addition to this range of the International Telecommunications Union's (ITU) standard schemes, which significantly outperforms all previous standard ITU codecs in robustness terms. The performance target of the 4 kbps ITU codec (ITU4) is also to maintain this impressive set of specifications. The family of codecs designed for various mobile radio systems – such as the 13 kbps regular pulse excited (RPE) scheme of the Global System of Mobile communications known as GSM, the 7.95 kbps IS-54, and the IS-95 Pan-American schemes, the 6.7 kbps Japanese digital cellular (JDC) and 3.45 kbps half-rate JDC arrangement (JDC/2) – exhibits slightly lower MOS values than the ITU codecs. Let us now consider the subjective quality of these schemes in a little more depth.

The 2.4 kbps US Department of Defence Federal Standard codec known as FS-1015 is the only vocoder in this group and it has a rather synthetic speech quality, associated with the lowest subjective assessment in the figure. The 64 kbps G.711 PCM codec and the G.726/G.727 adaptive differential PCM (ADPCM) schemes are waveform codecs. They exhibit a low implementational complexity associated with a modest bitrate economy. The remaining codecs belong to the so-called hybrid coding family and achieve significant bitrate economies at the cost of increased complexity and delay.

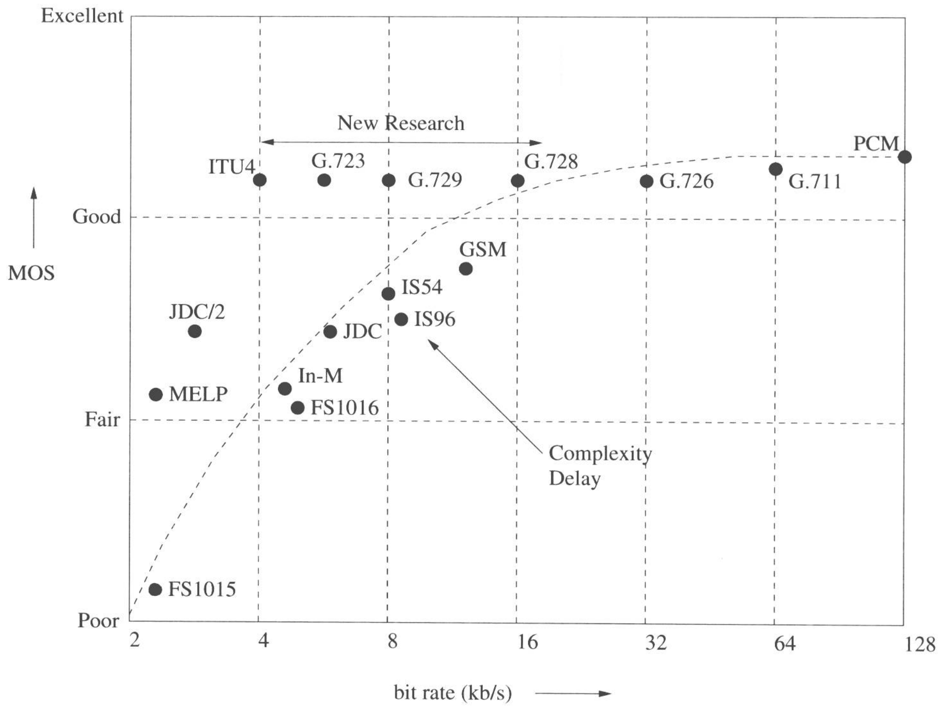


Figure 1: Subjective speech quality of various codecs [1] © IEEE, 1996.

Specifically, the 16 kbps G.728 backward-adaptive scheme maintains a similar speech quality to the 32 and 64 kbps waveform codecs, while also maintaining an impressively low, 2 ms delay. This scheme was standardised during the early 1990s. The similar quality, but significantly more robust 8 kbps G.729 codec was approved in March 1996 by the ITU. Its standardisation overlapped with the G.723.1 codec developments. The G.723.1 codec's 6.4 kbps mode maintains a speech quality similar to the G.711, G.726, G.727, G.728 and G.728 codecs, while its 5.3 kbps mode exhibits a speech quality similar to the cellular speech codecs of the late 1980s. The standardisation of a 4 kbps ITU scheme, which we refer to here as ITU4, is also a desirable design goal at the time of writing.

In parallel to the ITU's standardisation activities a range of speech coding standards have been proposed for regional cellular mobile systems. The standardisation of the 13 kbps RPE-long-term prediction (LTP) full-rate GSM (GSM-FR) codec dates back to the second half of the 1980s, representing the first standard hybrid codec. Its complexity is significantly lower than that of the more recent code excited linear predictive (CELP) based codecs. Observe in the figure that there is also a similar-rate enhanced full-rate GSM codec (GSM-EFR), which matches the speech quality of the G.729 and G.728 schemes. The original GSM-FR codec's development was followed a little later by the release of the 7.95 kbps vector sum excited linear predictive (VSELP) IS-54 American cellular standard. Due to advances in the field the 7.95 kbps IS-54 codec achieved a similar subjective speech quality to the 13 kbps GSM-FR scheme. The definition of the 6.7 kbps Japanese JDC VSELP codec was almost coincident with that of the IS-54 arrangement. This codec development was also followed by a half-rate standardisation process, leading to the 3.2 kbps pitch-synchroneous innovation CELP (PSI-CELP) scheme.

The IS-95 Pan-American code division multiple access (CDMA) system also has its own standardised CELP-based speech codec, which is a variable-rate scheme, supporting bitrates between 1.2 and 14.4 kbps, depending on the prevalent voice activity. The perceived speech quality of these cellular speech codecs contrived mainly during the late 1980s was found subjectively similar to each other under the perfect channel conditions of Figure 1. Lastly, the 5.6 kbps half-rate GSM codec (GSM-HR) also met its specification in terms of achieving a similar speech quality to the 13 kbps original GSM-FR arrangements, although at the cost of quadruple complexity and higher latency.

Recently, the advantages of intelligent multimode speech terminals (IMT), which can reconfigure themselves in a number of different bitrates, quality and robustness modes, attracted substantial research attention in the community, which led to the standardisation of the high-speed downlink packet access (HSDPA) mode of the 3G wireless systems. The HSDPA-style transceivers employ both adaptive modulation and adaptive channel coding, which result in a channel-quality dependent bitrate fluctuation, hence requiring reconfigurable multimode voice and audio codecs, such as the advanced multirate codec, referred to as the AMR scheme. Following the standardisation of the narrowband AMR codec, the wideband AMR scheme, referred to as the AMR-WB arrangement and encoding the 0–7 kHz band, was also developed, which will also be characterised in this book. Finally, the most recent AMR codec, namely the so-called AMR-WB+ scheme, will also be the subject of our discussions.

Recent research on sub-2.4 kbps speech codecs is also covered extensively in this book, where the aspects of auditory masking become more dominant. Finally, since the classic G.722 sub-band-adaptive differential pulse code modulation (ADPCM) based wideband codec has become obsolete in the light of exciting new developments in compression, the most recent trend is to consider wideband speech and audio codecs, providing substantially enhanced speech quality. Motivated by early seminal work on transform-domain or frequency-domain based compression by Noll and his colleagues, in this field the wideband G.721.1 codec – which can be programmed to operate between 10 kbps and 32 kbps and hence lends itself to employment in HSDPA-style near-instantaneously adaptive wireless communicators – is the most attractive candidate. This codec is portrayed in the context of a sophisticated burst-by-burst adaptive wideband turbo-coded orthogonal frequency division multiplex (OFDM) IMT in this book. This scheme is also capable of transmitting high-quality audio signals, behaving essentially as a high-quality waveform codec.

Milestones in Speech Coding History

Over the years a range of excellent monographs and text books have been published, characterising the state-of-the-art at its various stages of development and constituting significant milestones. The first major development in the history of speech compression can be considered to be the invention of the vocoder, dating back to as early as 1939. Delta modulation was contrived in 1952 and later it became well established following Steele's monograph on the topic in 1975 [3]. PCM was first documented in detail in Cattermole's classic contribution in 1969 [4]. However, it was realised in 1967 that predictive coding provides advantages over memoryless coding techniques, such as PCM. Predictive techniques were analysed in depth by Markel and Gray in their 1976 classic treatise [5]. This was shortly followed by the often cited reference [6] by Rabiner and Schafer. Also, Lindblom and Ohman contributed a book in 1979 on speech communication research [7].

The foundations of auditory theory were laid down as early as 1970 by Tobias [8], but these principles were not exploited to their full potential until the invention of the analysis-by-synthesis (AbS) codecs, which were heralded by Atal's multi-pulse excited codec in the early 1980s [9]. The waveform coding of speech and video signals has been comprehensively documented by Jayant and Noll in their 1984 monograph [10]. During the 1980s the speech codec developments were fuelled by the emergence of mobile radio systems, where spectrum was a scarce resource, potentially doubling the number of subscribers and hence the revenue, if the bitrate could be halved.

The RPE principle – as a relatively low-complexity AbS technique – was proposed by Kroon, Deprettere and Sluyter in 1986 [11], which was followed by further research conducted by Vary [12,13] and his colleagues at PKI in Germany and IBM in France, leading to the 13 kbps Pan-European GSM codec. This was the first standardised AbS speech codec, which also employed LTP, recognising the important role the pitch determination plays in efficient speech compression [14,15]. It was in this era, when Atal and Schroeder invented the code excited linear predictive (CELP) principle [16], leading to perhaps the most productive period in the history of speech coding during the 1980s. Some of these developments were also summarised, for example, by O'Shaughnessy [17], Papamichalis [18] and Deller, Proakis and Hansen [19].

It was during this era that the importance of speech perception and acoustic phonetics was duly recognised, for example, in the monograph by Lieberman and Blumstein [20]. A range of associated speech quality measures were summarised by Quackenbush, Barnwell III and Clements [21]. Nearly concomitantly Furui also published a book related to speech processing [22]. This period witnessed the appearance of many of the speech codecs seen in Figure 1, which found applications in the emerging global mobile radio systems, such as IS-54, JDC, etc. These codecs were typically associated with source-sensitivity matched error protection, where, for example, Steele, Sundberg and Wong [23–26] have provided early insights on the topic. Further sophisticated solutions were suggested, for example, by Hagenauer [27].

Both the narrowband and wideband AMR, as well as the AMR-WB+ codecs [28,29] are capable of adaptively adjusting their bitrate. This also allows the user to adjust the ratio between the speech bitrate and the channel coding bitrate constituting the error protection oriented redundancy according to the prevalent near-instantaneous channel conditions in HSDPA-style transceivers. When the channel quality is inferior, the speech encoder operates at low bitrates, thus accommodating more powerful forward error control within the total bitrate budget. By contrast, under high-quality channel conditions the speech encoder may benefit from using the total bitrate budget, yielding high speech quality, since in this high-rate case low redundancy error protection is sufficient. Thus, the AMR concept allows the system to operate in an error-resilient mode under poor channel conditions, while benefitting from a better speech quality under good channel conditions. Hence, the source coding scheme must be designed for seamless switching between rates available without annoying artifacts.

Overview of MPEG-4 Audio

The definition of the MPEG-4 audio standard was the culmination of the 60-year research conducted by the global research community, as portrayed in Figure 3, which will be detailed throughout out discussions in the book. The Moving Picture Experts Group (MPEG) was first established by the International Standard Organisation (ISO) in 1988 with the aim of developing a full audio-visual coding standard referred to as MPEG-1 [30–32]. The audio-related section MPEG-1 was designed to encode digital stereo sound at a total bitrate of 1.4 to 1.5 Mbps – depending on the sampling frequency, which was 44.1 kHz or 48 kHz – down to a few hundred kilobits per second [33]. The MPEG-1 standard is structured in layers, from Layer I to III. The higher layers achieve a higher compression ratio, albeit at an increased complexity. Layer I achieves perceptual transparency, i.e. subjective equivalence with the uncompressed original audio signal at 384 kbps, while Layer II and III achieve a similar subjective quality at 256 kbps and 192 kbps, respectively [34–38].

MPEG-1 was approved in November 1992 and its Layer I and II versions were immediately employed in practical systems. However, the MPEG Audio Layer III, MP3 for short only became a practical reality a few years later, when multimedia PCs were introduced having improved processing capabilities and the emerging Internet sparked off a proliferation of MP3 compressed teletraffic. This changed the face of the music world and the distribution of music. The MPEG-2 backward compatible audio standard was approved in 1994 [39], providing an improved technology that would allow those who had already launched MPEG-1 stereo audio services to upgrade their system to multichannel mode, optionally also supporting a higher number of channels at a higher compression ratio. Potential applications of the multichannel mode are in the field of quadraphonic music distribution or cinemas. Furthermore, lower sampling frequencies were also incorporated, which include 16, 22.05, 24, 32, 44.1 and 48 kHz [39]. Concurrently, MPEG commenced research into even higher-compression schemes, relinquishing the backward compatibility requirement, which resulted in the MPEG-2 advanced audio coding standard (AAC) standard in 1997 [40]. This provides those who are not constrained by legacy systems to benefit from an improved multichannel coding scheme. In conjunction with AAC, it is possible to achieve perceptual transparent stereo quality at 128 kbps and transparent multichannel quality at 320 kbps; for example in cinema-type applications.

The MPEG-4 audio recommendation is the latest standard completed in 1999 [41–45], which offers, in addition to compression, further unique features that will allow users to interact with the information content at a significant higher level of sophistication than is possible today. In terms of compression, MPEG-4 supports the encoding of speech signals at bitrates from 2 kbps up to 24 kbps. For coding of general audio, ranging from very low bitrates up to high quality, a wide range of bitrates and bandwidths are supported, ranging from a bitrate of 8 kbps and a bandwidth below 4 kHz to broadcast quality audio, including monaural representations up to multichannel configuration.

The MPEG-4 audio codec includes coding tools from several different encoding families, covering parametric speech coding, CELP-based speech coding and time/frequency (T/F)

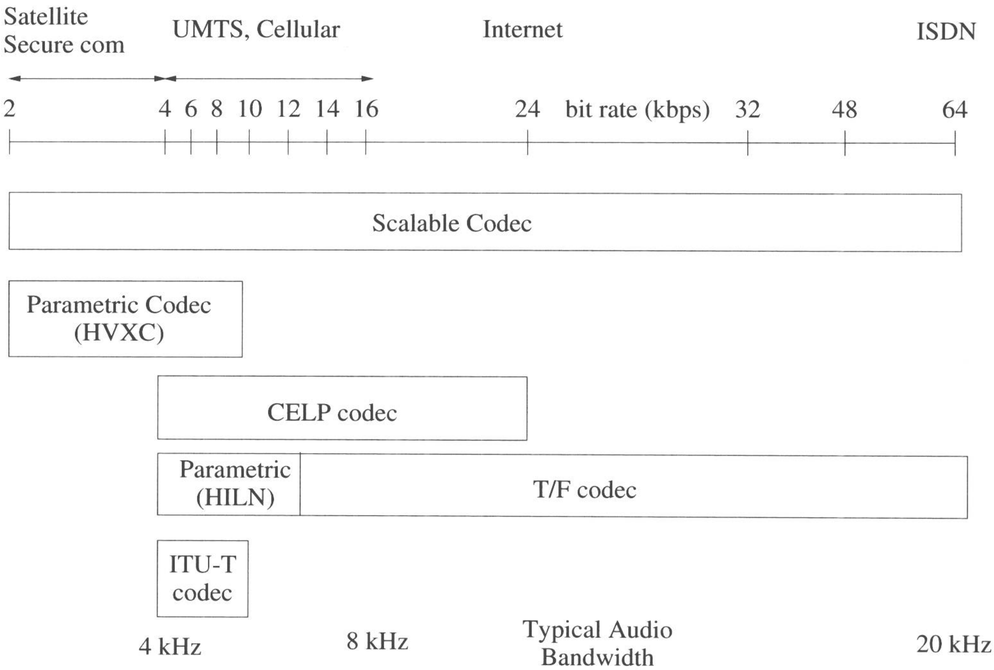


Figure 2: MPEG-4 framework [41].

audio coding, which are characterised in Figure 2. It can be observed that a parametric coding scheme, namely Harmonic Vector eXcitation Coding (HVXC) was selected for covering the bitrate range from 2 to 4 kbps. For bitrates between 4 and 24 kbps, a CELP-coding scheme was chosen for encoding narrowband and wideband speech signals. For encoding general audio signals at bitrates between 8 and 64 kbps, a T/F coding scheme based on the MPEG-2 AAC standard [40] endowed with additional tools is used. Here, a combination of different techniques was established, because it was found that maintaining the required performance for representing speech and music signals at all desired bitrates cannot be achieved by selecting a single coding architecture. A major objective of the MPEG-4 audio encoder is to reduce the bitrate, while maintaining a sufficiently high flexibility in terms of bitrate selection. The MPEG-4 codec also offers other new functionalities, which include bitrate scalability, object-based of a specific audio passage for example, where a distinct 'object' may be defined as a passage played by a certain instrument coding, as well as an increased robustness against transmission errors and supporting special audio effects.

MPEG-4 consists of Versions 1 and 2. Version 1 [41] contains the main body of the standard, while Version 2 [46] provides further enhancement tools and functionalities, that includes the issues of increasing the robustness against transmission errors and error protection, low-delay audio coding, finely grained bitrate scalability using the Bit-Sliced Arithmetic Coding (BSAC) tool, the employment of parametric audio coding, using the CELP-based silence compression tool and the 4 kbps extended variable bitrate mode of the HVXC tool. Due to the vast amount of information contained in the MPEG-4 standard, we will only consider some of its audio compression components, which include the coding of natural speech and audio signals. Readers who are specifically interested in text-to-speech synthesis or synthetic audio issues are referred to the MPEG-4 standard [41] and to the contributions by Scheirer *et al.* [47, 48] for further information. Most of the material in Chapter 10 will be based on an amalgam of [34–38, 40, 41, 43, 44, 46, 49]. In this chapter, the operations of each component of the MPEG-4 audio component will be highlighted in greater detail. As an application example, we will employ the transform-domain weighted interleaved vector quantisation (TWINVQ) coding tool, which is one of the MPEG-4 audio codecs in the context of a wireless audio transceiver in conjunction with space–time coding [50] and various quadrature amplitude modulation (QAM) schemes [51]. The audio transceiver is introduced in Section 10.5 and its performance is discussed in Section 10.5.6.

Motivation and Outline of this Book

During the early 1990s, Atal, Cuperman and Gersho [52] edited prestigious contributions on speech compression. Also, Ince [53] contributed a book in 1992 related to the topic. Anderson and Mohan co-authored a monograph on source and channel coding in 1993 [54]. Research-oriented developments were then consolidated in Kondoz' excellent monograph in 1994 [55] and in the multi-authored contribution edited by Kleijn and Paliwal [56] in 1995. The most recent addition to the above range of contributions is the second edition of O'Shaughnessy well-referenced book cited above. However, at the time of writing no book spans the entire history of speech and audio compression, which is the goal of this volume.

Against this backcloth, this book endeavours to review the recent history of speech compression and communications in the era of wireless turbo-transceivers and joint source/channel coding. We attempt to provide the reader with a historical perspective, commencing with a rudimentary introduction to communications aspects, since throughout this book we illustrate the expected performance of the various speech codecs studied also in the context of jointly optimised wireless transceivers.

This book contains four parts. Parts I and II cover classic background material on speech signals, predictive waveform codecs and analysis-by-synthesis codecs as well as the entire speech and audio coding standardisation scene. The bulk of the book is contained in the research-oriented Parts III and IV, covering both standardised and proprietary speech codecs – including the most recent AMR-WB+ and the MPEG-4 audio codecs, as well as cutting-edge wireless turbo transceivers.

Specifically, Chapters 1 and 2 of Part I provide a rudimentary introduction to speech signals, classic waveform coding as well as predictive coding, respectively, quantifying the overall performance of the various speech codecs, in order to render our treatment of the topics as self-contained and all-encompassing as possible.

Part II of this book is centred around AbS based coding, reviewing the classic principles in Chapter 3 as well as both narrow and wideband spectral envelope quantisation in Chapter 4. RPE and CELP coding are the topic of Chapters 5 and 6, which are followed by a detailed chapter on the entire plethora of existing forward-adaptive standardised CELP codecs in Chapter 7 and on their associated source-sensitivity matched channel coding schemes. The subject of Chapter 8 is both proprietary and standard backward-adaptive CELP codecs,

Algorithms/Techniques	Timeline	Standards/Commercial Codecs
Fletcher: Auditory patterns [81]	1940	
Zwicker, Greenwood: Critical bands [82,83]	1961	
Scharf, Hellman: Masking effects [84,85]	1970	
Schroeder: Spread of masking [86]	1979	
Nussbaumer: Pseudo-Quadrature Mirror Filter [87]	1981	
Rothweiler: Polyphase Quadrature Filter [88]	1983	
Princen: Time Domain Aliasing Cancellation [89]	1986	
	1987	
Johnston: Perceptual Transform Coding [90]	1988	AT&T: Perceptual Audio Coder (PAC) [102]
Mahieux: backward adaptive prediction [91] Edler: Window switching strategy [92] Johnston: M/S stereo coding [93]	1989	CNET codec [91]
Malvar: Modified Discrete Cosine Transform [94]	1990	
	1991	Dolby AC-2 [103]
	1992	MPEG-1 Audio finalized [104] Dolby AC-3 [103]
	1993	Sony: MiniDisc: Adaptive Transform Acoustic Coding (ATRAC) [105]
Herre: Intensity Stereo Coding [95]	1994	Philips: Digital Compact Cassette (DCC) [106] MPEG-2 backward compatible [107]
Iwakami: TWINVQ [96] Herre & Johnston: Temporal Noise Shaping [97]	1995	NTT: Transform-domain Weighted Interleaved Vector Quantization (TWINVQ) [96,108]
Park: Bit-Sliced Arithmetic Coding (BSAC) [98]	1997	MPEG-2 Advanced Audio Coding (AAC) [109]
Purnhagen: Parametric Audio Coding [99]	1998	
Levine & Smith, Verma & Ming: Sinusoidal+Transients+Noise coding [100,101]	1999	MPEG-4 Version 1 & 2 finalized [110,111]

Figure 3: Important milestones in the development of perceptual audio coding.

which is concluded with a system design example based on a low-delay, multimode wireless transceiver.

The research-oriented Part III of this book is dedicated to a range of standard and proprietary wideband coding techniques and wireless systems. As an introduction to the wideband coding scene, in Chapter 9 the classic sub-band-based G.722 wideband codec is reviewed first, leading to the discussion of numerous low-rate wideband voice and audio codecs. Chapter 9 also contains diverse sophisticated wireless voice- and audio-system design examples, including a turbo-coded OFDM wideband audio system design study. This is followed by a wideband voice transceiver application example using the AMR-WB codec, a source-sensitivity matched Irregular Convolutional Code (IRCC) and extrinsic information transfer (EXIT) charts for achieving a near-capacity system performance. Chapter 9 is concluded with the portrayal of the AMR-WB+ codec. In Chapter 10 of Part III we detail the principles behind the MPEG-4 codec and comparatively studied the performance of the MPEG-4 and AMR-WB audio/speech codecs combined with various sophisticated wireless transceivers. Amongst others, a jointly optimised source-coding, outer unequal protection non-systematic convolutional (NSC) channel-coding, inner trellis coded modulation (TCM) and spatial diversity aided space–time trellis coded (STTC) turbo transceiver investigated. The employment of TCM provided further error protection without expanding the bandwidth of the system and by utilising STTC spatial diversity was attained, which rendered the error statistics experienced pseudo-random, as required by the TCM scheme, since it was designed for Gaussian channels inflicting randomly dispersed channel errors. Finally, the performance of the STTC-TCM-2NSC scheme was enhanced with the advent of an efficient iterative joint decoding structure.

Chapters 11–17 of Part IV are all dedicated to sub-4 kbps codecs and their wireless transceivers, while Chapter 18 is devoted to speech quality evaluation techniques as well as to a rudimentary comparison of various speech codecs and transceivers. The last chapter of the book is on VoIP.

This book is naturally limited in terms of its coverage of these aspects, simply owing to space limitations. We endeavoured, however, to provide the reader with a broad range of application examples, which are pertinent to a range of typical wireless transmission scenarios.

Our hope is that this book offers you – the reader – a range of interesting topics, portraying the current state-of-the-art in the associated enabling technologies. In simple terms, finding a specific solution to a voice communications problem has to be based on a compromise in terms of the inherently contradictory constraints of speech quality, bitrate, delay, robustness against channel errors, and the associated implementational complexity. Analysing these trade-offs and proposing a range of attractive solutions to various voice communications problems is the basic aim of this book.

Again, it is our hope that this book underlines the range of contradictory system design trade-offs in an unbiased fashion and that you will be able to glean information from it, in order to solve your own particular wireless voice communications problem, but most of all that you will find it an enjoyable and relatively effortless reading, providing you – the reader – with intellectual stimulation.

Lajos Hanzo
Clare Somerville
Jason Woodard