Joint Source-Channel Decoding

A Cross-Layer Perspective with Applications in Video Broadcasting over Mobile and Wireless Networks

Pierre Duhamel
Michel Kieffer

Joint Source-Channel Decoding

A Cross-Layer Perspective with Applications in Video Broadcasting over Mobile and Wireless Networks

Pierre Duhamel
Michel Kieffer

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier

AP

Academic Press is an imprint of Elsevier
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA

First edition 2010

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress.

ISBN-13: 978-0-12-374449-4

For information on all Academic Press publications
visit our Web site at *www.elsevierdirect.com*

Typeset by: diacriTech, India

Printed and bound in the UK
10 11 12 13 14 10 9 8 7 6 5 4 3 2 1

Joint Source-Channel Decoding

EURASIP and Academic Press Series in Signal and Image Processing

Series Advisor

Sergios Theodoridis

Titles in the series:

Duhamel and Kieffer

Joint Source-Channel Decoding, ISBN 9780123744494

Thiran, Marqués and Bourlard

Multimodal Signal Processing, ISBN 9780123748256

Preface and Acknowledgements

Il ne faut pas toujours tellement épuiser un sujet qu'on ne laisse rien à faire au lecteur. Il ne s'agit pas de faire lire mais de faire penser.
Montesquieu (L'esprit des lois)

(approximate) English translation (and in short):
One should not always exhaust a subject in such a way that this leaves nothing to do to the reader. This is not to read but to think.

Joint source-channel coding/decoding is a rather specialized term for designating techniques allowing to transmit efficiently audiovisual contents through wireless channels by means of possibly unconventional methods. Such a simple goal turns out to be quite complex since it involves many different techniques (namely, source coding, networking, and transmission techniques) usually mastered by different teams.

Therefore, this book does not intend to address with full accuracy all these aspects, but rather tries to give a precise idea of the basics of each and to provide a state of the art of the "joint source/network/channel decoding" results obtained on video signals. Many examples illustrate the corresponding concepts and algorithms. It is essentially based on a tutorial by the same authors, which was presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing in 2007. The joint source-channel coding (JSCC) aspect is outlined only at the end of the book.

Being a topic with such a wide scope, we were obviously helped and stimulated by many people. Our only hope is that we will not forget too many of them in the list below.

Very special thanks are due to several colleagues: Olivier Rioul, who introduced us to the topic of JSCC, before we even thought of JSCD, Pierre Siohan, who always showed a strong interest in this topic (and led one of the corresponding projects), and Christine Guillemot and her team, with whom the neologism "coopetition" seems to be the most appropriate for characterizing our stimulating cooperation/competition. Catherine Lamy-Bergot and Kave Salamatian also took part in many fruitful discussions on the topic. Parts of this book were proofread by colleagues who did a very detailed job, many thanks to Aurelia Fraysse, Jean-Benoist Leger, Pablo Piantanida, Phil Schniter, and Claudio Weidmann.

Obviously, when working on this topic, we were helped by many PhD students, namely, Barham Zahirazami, Avi Gabay, Hang Nguyen, Rachel Chiang, Chang-Ming Lee, Galina Sabeva, Salma Ben-Jamaa, Cedric Marin, Usman Ali, and Zied Jaoua who worked directly on the topics covered by this book. All of them were funded by either industry grants or collaborative contracts funded by the French government. Other PhD students, namely, José Saavedra and Olivier Derrien, were also of great help for some parts of the book.

In fact, we were very lucky to obtain funding for such a long-term research topic, almost continuously since 1997. Special thanks are thus due to Gerard Eude, from France Telecom R&D, who initiated these fundings, followed by CNES (the French Space Agency, with Catherine Lambert-Nebout as an interface), and Alcatel-Lucent, a long-term contributor to this work, with an initial strong commitment of Denis Rouffet. CNES (with Gaël Scot as contact person, and Jerôme Lacan as coworker) also funded a cross-layer study on JSCD, which demonstrated to us that both aspects were two sides of the same coin.

Part of the material described in this book was funded by the French Ministry of Research and Industry by means of collaborative contracts named COSOCATI, DIVINE, VIP, DITEMOI, and TVMSL.

Finally, most of this work benefited from contacts established by means of two Networks of Excellence of the European Commission, NEWCOM and NEWCOM++, scientifically led in a friendly and efficient manner by Sergio Benedetto. We also benefited from fundings through these NoEs and helpful collaborative work with Lajos Hanzo, Michael Ng, and Morten Holm Larsen.

The final thanks will go to the people who showed the highest patience during the writing of this book: our families, who did not sign to write the book, but had to undergo the consequences... and our editor, Tim Pitts, whose patience and sense of humor were combined with sufficient pressure so that the book finally came to an end without too much delay.

More materials (Appendices X, Y and Z) can be found at website: http://www.elsevierdirect.com/companions/9780123744494.

Contents

Chapter 1

Introduction: Context

This book addresses a question of growing importance: How can one improve the perceived quality of video communicated over a wireless network? For successful communication of multimedia data, especially video, the network must behave as if it is predictable and stable. Wireless networks are not known for predictability and stability, but rather variability and scarcity of resources. Oversizing, the usual way of coping with variability, is not practical in wireless systems due to the scarcity of resources. Yet, the number of applications for video broadcasting over wireless channels keeps increasing. Therefore, there is a growing need for techniques that provide sufficient received video quality without overspending the scarce resources of the wireless network.

In the classical approach to network design, the networking task is partitioned into distinct layers, where each layer, assuming that the lower layers behave perfectly, attempts to provide perfect information to the upper layers. In recent years, the so-called joint approaches have been proposed for improved performance and better use of resources. In joint approaches, the network layers become more transparent so that the information previously available at only one layer is accessible to other layers. The downside to this ambitious approach is a loss of the architectural coherence that was the primary driving force behind the use of decoupled layers.

Wireless network designs, thus, fall between two extreme points: a generic layered architecture, which would be safe but not very efficient at resource utilization, and an optimized architecture (i.e., optimal in both resource utilization and performance), which would be dedicated to a single application. Finding joint mechanisms that are compatible with the use of layered, generic architectures is a serious challenge.

Given the complexity of the task, the joint mechanisms usually proposed are quite limited. Most techniques adapt the parameters used in source coding, channel coding, and other procedures based on a limited set of information provided by other layers and thus, do not reach their full potential. While these so-called cross-layer approaches have made some impact on wireless network performance, they are not truly joint schemes and constitute a first step toward a global optimization of the network. In fact, there seems to be a large difficulty in obtaining fully joint techniques compatible with generic network architectures.

This book intends to fill the gap by developing – in some detail – tools that allow the implementation of fully joint techniques for robust reception of multimedia transmitted over wireless/Internet Protocol (IP) networks. Since these techniques require the cooperation of many layers, each with its own peculiar characteristics; this book covers a wide variety of topics from source coding to channel coding to networking.

While there are many books that focus on any one of these topics (e.g., source coding, channel coding, wireless communications, or wireless networking), the readerships of these books are often very different. For example, the community that works on source coding does not overlap too much with the one that works on the so-called physical layer. So, years ago, a tool known as joint source-channel coding (JSCC) was proposed to address optimal source transmission for applications in which Shannon's separation principle does not hold. However, the problem was considered as a curiosity by many practitioners, and even if a (relatively small) number of articles have appeared regularly on this topic, the impacts have not been widespread. The main drawback of JSCC is that, since it is not compatible with any existing standards, widespread implementation would require significant time and effort, even if the performance improvements were known to be potentially worthwhile. Motivated by the practical challenges of JSCC, an interest in joint source-channel decoding (JSCD) emerged, whereby the goal is to extract the best performance out of the received signal without considering changes to the transmitted signal. Clearly, the ability to use JSCD with existing standards makes it potentially very practical. Therefore, there is a good possibility that JSCD tools could become useful in the near future, provided that all the different layers can be incorporated. For example, many JSCD articles treat the network layers as transparent, which would clearly be a drawback for practical implementation. At the same time, cross-layer communication protocols have become a great interest to some who, motivated by multimedia applications, have worked on jointly adapting the characteristics of each layer to the type of signal that was transmitted. The aim of this book is precisely to provide a cross-layer perspective on JSCC/JSCD tools so that their applicability is increased; on one hand, JSCD tools become more practical when network layers are not treated as fully transparent, and on the other hand, the cross-layer approach benefits from additional efficient tools.

Though this book necessarily treats topics from a wide range of fields (e.g., source coding, channel coding, and networking), it attempts to be as self-contained as possible.

1.1 MULTIMEDIA WIRELESS: THE NEED FOR NEW TOOLS

1.1.1 The Classical Approach to Multimedia Wireless

Many years ago, when communication systems were first designed, they were targeted toward a specific application. For example, the wired phone system was principally designed to carry speech signals, as were the first generations of wireless phone systems, while the Internet was designed to carry data files. From that starting point, these systems have evolved greatly and now support a much wider variety of applications, using (almost) the same channels and protocols. For example, fax was added to the wired phone system, short message services (SMS) and multimedia services (MMS) were added to wireless phone systems, and Voice over IP (VoIP) was added to the Internet. Furthermore, the systems have been integrated to the point that, now, many phone calls go (at least partially) through the Internet without users being notified or even able to tell the difference, and many e-mails are typed and/or read on wireless mobile phones.

Moreover, the bitrate available to most systems has grown so high that they now support very demanding signals. For example, high-definition television (HDTV) goes through wired phone lines, via asymmetric digital subscriber line (ADSL), and video is routinely transmitted through wireless channels. As discussed below, these changes have led to pronounced shifts in both the usage and the design of these systems.

Changes in System Design

The dominant approach to the design of wireless multimedia systems is the layered approach, where the network layers are designed to function independently from each other. Each layer is dedicated to a specific networking task, and the layers are designed regardless of the signal types carried by the network.

This layered approach is motivated to some extent by Shannon's separation principle for source-channel coding, which (loosely speaking) established that, under certain conditions, the best joint source/channel coder is the concatenation of the best source coder with the best channel coder. In such situations, the source coder generates the lowest-rate bitstream given an allowed distortion – free of redundancy, whereas the channel coder generates the lowest-rate bitstream enabling the correction of "almost" all channel-induced errors – via judicious use of redundancy. Thus, the optimal source coding is invariant to the channel characteristics, whereas optimal channel coding is invariant to the source characteristics. However, this separation principle (described in more detail in Chapter 2) holds only under certain not-so-practical conditions, such as in the limit of infinite code-word length. Implementations that attempt to satisfy these conditions directly would involve very high complexity and possibly very large transmission delays.

Changes in System Usage

Predicting the way that wireless multimedia systems will be used turns out to be very difficult. In the early 1970s, during the time of analog telephones, there were already proposals for videophones in various parts of the world. At that time, cost was a significant problem. Several years later, in the early 1990s, digital techniques had evolved, and the first studies of video compression (again targeted at videophones, i.e., visiophony) resulted in the widely accepted H.263 standard, published in 1996. It took some time for this technology to be viable on widely deployed terminals, but at that time, visiophony did not seem to be a driving need. Only recently did some "killer applications" for mobile networks emerge: video database consultation and video broadcasting. While these uses of wireless networks appear to be linked primarily to the ways that mobile phones are now being used (e.g., more like a PC for surfing the Internet and downloading files, images, and videos), they show a clear shift in the way that networks are being used: the resources required for transmitting a 3-minute video clip are far beyond those required to support voice communication for the same duration. Even downloading audio files is about 10 times more costly (in terms of bitrate) than communicating speech. Since multimedia traffic is increasing at a much faster rate than speech traffic, one is well justified to concentrate on the multimedia aspects of mobile devices.

As a result of these trends, it may well happen that demands for increased bandwidth, in conjunction with the use of limited-efficiency layered networks, will overload the global network. While many actions have been taken to prevent this occurrence (e.g., the addition of broadcast mode in some standards), new tools as JSCC/JSCD and cross-layer optimization could be of great help.

1.1.2 A New Approach to Multimedia Wireless

The claim that current networks are implemented using fully independent layers is really an oversimplification. For example, the second generation European mobile telephone system (GSM) residual bit error rate (BER) is tuned so that it is compatible with the BER tolerated by GSM's speech coder. This, in fact, is already a form of cross-layer optimization in which the behavior of the physical (PHY) layer is matched to the needs of the application layer. Clearly, the tuning would be different for other applications and signal types (e.g., video).

The primary factors motivating the shift from fully independent layers to closely coupled layers are the high variability of wireless channels and the need for increased robustness and graceful degradation. If a network must support a wide variety of channels and users, then designing the network to work perfectly on all of them amounts (for the most part) to designing the network to satisfy the worst channel/user. In this case, if one user experiences a very bad channel, then that user would consume a disproportionate amount of system resources.

In addition, it should be remembered that the wireless multimedia networks are composed of many links, some wireless (e.g., to the terminals) and some wired. The means of protecting the signal is somewhat different over these two links due to the differing natures of their impairments. For the wired portion, the network adds headers (containing, e.g., the MAC address of the user and the number assigned to each packet) and cyclic redundancy checks (CRCs), which are used to check the integrity of the received packet. Because of the reliability of the wired channel, it is feasible to implement protection using an acknowledgment (ACK/NACK)-based procedure known as Automatic Repeat reQuest (ARQ): a packet is retransmitted if, upon reception, the CRC does not match the part that it protects (e.g., the payload, header, or both). Clearly, ARQ-based error correction requires feedback and induces large delays (though hopefully on only a few packets). If feedback was somehow unavailable (e.g., in the broadcast scenario), then the CRC would at least indicate which packets were not correctly received.

The amount of delay that is tolerable, as well as the channel conditions themselves, can vary widely across different applications.

Finally, practical communication links are rarely dedicated entirely to a single user, and a given user rarely communicates through only a single link. Rather, each user's signal is segmented into packets which are scheduled, along with the packets of many other users, through available links. This segmentation of packets has often been overlooked in the design of robust reception schemes, even though it can have a large (positive or negative) impact on those strategies. This topic will be addressed in the second part of the book, i.e., Chapters 7–9, describing the corresponding tools that are similar to the ones used in more traditional JSCD (explained in the first part).

To make these notions more concrete, a few examples of applications that might benefit from the tools discussed in this book are described, and their essential characteristics are analyzed.

1.2 EXAMPLE APPLICATIONS

Over the past several years, the authors have worked with engineers, network operators, and academic colleagues (all of them gratefully acknowledged) on various wireless multimedia system design projects. Some of these projects are described here and linked to later chapters of the book. These examples are useful in bridging the abstract models used later in the book with practical real-world problems.

1.2.1 Example 1: Visiophony Over IP (RNRT VIP)

This project was concerned with the classical scenario of a video communication through wireless terminal links that are connected by a wired central link that

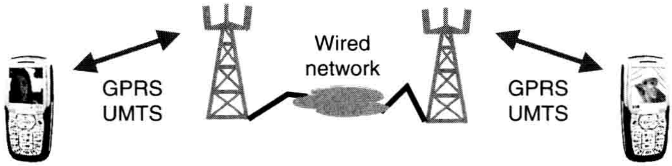

FIGURE 1.1 Wireless visiophony over IP.

uses the IP, see Figure 1.1. Admittedly, the model assumed in the project was abstract, and many implementation details were never considered. However, the essential constraints resulting from the visiophony application were (1) that the maximum delay induced by coding and transmission was at most ≈ 0.5s and (2) that the reconstructed speech/video signal was of sufficiently high quality.

The proposed solution leveraged the fact that the error mechanism on the wired link is different than that on the wireless link. For example, the wired link can introduce packet losses, which – in terms of transmission – are equivalent to erasures. While the delay constraint makes the retransmission of lost packets difficult, the knowledge of which packets are lost can be used for error concealment; portions of the stream received without error can be interpolated to "fill the holes." In contrast, the wireless link must contend with intersymbol interference (ISI) and noise, the mitigation of which can be accomplished using JSCD. Since it was assumed that the wireless link does not support retransmission, the decoder must be able to display an image on the screen, whatever the output of the channel. But even in this situation, CRC checks can be exploited by the PHY layer: the packets that can be correctly received using a simple "hard decision" procedure (as indicated by the CRC) do not need the application of a more complex JSCD procedure. Packets that do not pass this check can be processed by a robust decoder, after which CRC-based checks would be infeasible.

While the mechanisms described above are somewhat simple and ignore many practical aspects of implementation, they already go beyond what has been proposed in many JSCD studies, where, e.g., erasures were not taken into account. The primary aim of this study was to quantify the potential improvements of "classical" JSCD. (See Figure 1.2.)

Example 1: Summary

- Mixed wired/wireless links facilitate error concealment.
- Strong delay constraints limit the possibility for ARQ.
- CRCs can be used to determine which packets require complex JSCD.

1.2.2 Example 2: Video Broadcasting (CNES SDMB, Alcatel TVMSL)

The second example pertains to a project launched by Alcatel-Lucent in 2006, partially funded by the "Agence de l'Innovation Industrielle" (AII) of the French

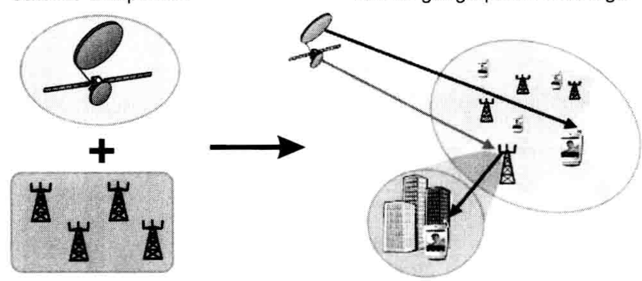

FIGURE 1.2 Video broadcasting using a mixture of satellite and terrestrial links.

government, where the goal was to provide (relatively) low-cost video broadcasting to mobiles. The project was motivated by the observation that the first users of 3G were using video consultation much more than anticipated. In the proposed solution, mobiles were addressed by satellites when possible or terrestrial relays when satellite reception was poor, see Figure 1.2. While, in this project, most of the effort was devoted to implementation, some effort was made to evaluate the impact of JSCD.

At first glance, this application does not seem to impose a delay constraint; since there is no need to support a comfortable conversation between two users, there does not seem to be a drawback to a several-second delay between source and destination. However, it turns out that users do want the ability to switch from one channel to another without waiting too long. Thus, the application does in fact impose a delay constraint, though perhaps not as strict as the one in the previous example. For this project, the DVS-S/H standard was assumed, which allowed the introduction of MAC-layer redundancy but required cross-layer tuning. Since a broadcast scenario was considered, feedback was not feasible, and the system was designed to maximize robustness and coverage. Finally, a wide range of terminal mobility (i.e., slow to fast) was allowed. Since mobility greatly impacts the fading characteristics of the wireless channel, it can greatly impact performance in terms of video quality.

Example 2: Summary

- Primarily, wireless links
- Mild delay constraints, e.g., up to several seconds
- No possibility of feedback
- Wide range of terminal mobility

1.2.3 Example 3: Multimedia Multicasting (RNRT DIVINE)

The application envisioned here is a museum in which supplemental information on each exhibit, in the form of multimedia files, is served to visitors' mobile

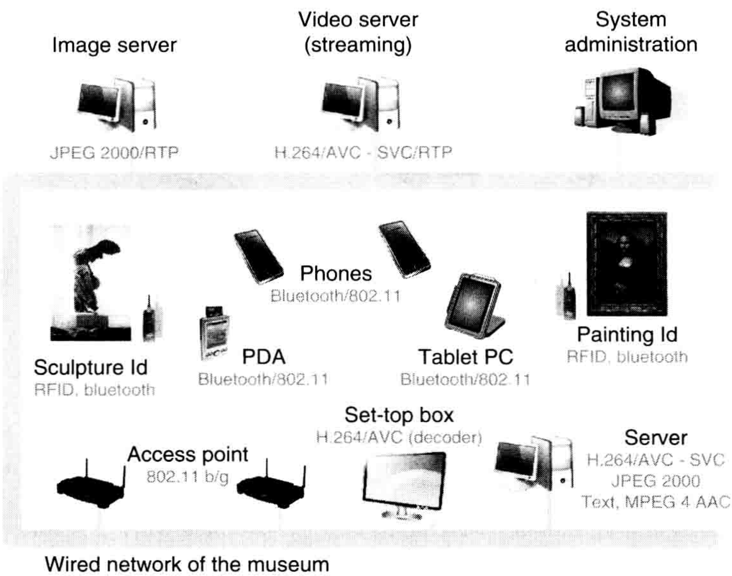

FIGURE 1.3 Multicasting on wireless links.

phones or Personal digital assistants (PDAs) via Wi-Fi. By using their own devices, visitors can avoid the need to rent dedicated devices, which often have fixed content (e.g., prerecorded audio) and offer little flexibility. Figure 1.3 illustrates some possibilities for such a situation, e.g., the reference number of the exhibit could be obtained wirelessly, allowing access to audio, images, or video, which provide useful information on the exhibit.

Besides providing a practical service, this project provides an interesting example of a scenario that falls somewhere between point-to-point and broadcast communication, because of the fact that several groups of visitors might want to download the same files at about the same time. In fact, by grouping these users and assigning a single channel to each group, the problem can be considered as a multicasting one. Here, a judicious use of delay can be used to synchronize users within a group. Relative to point-to-point communication, multicasting uses less channels, and relative to full broadcast, multicasting allows the transmission of only necessary files. Furthermore, the use of feedback (e.g., ARQ) is feasible, since the number of users per channel should not be very large (and, furthermore, can be controlled). In fact, choosing the best number of users per group is an interesting problem, considering the effects of power allocation and ARQ within a group.

Example 3: Summary

- Mixed wired wireless channels (though essentially wireless)