Preface, Motivation and The Speech Coding Scene

In the era of third-generation (3G) wireless personal communications standards, despite the emergence of broad-band access network standard proposals, the most important mobile radio services are still based on voice communications. Even when the predicted surge of wireless data and Internet services becomes a reality, voice will remain the most natural means of human communication, although it may be delivered via the Internet, predominantly after compression.

This book is dedicated mainly to voice compression issues. Error resilience, coding delay, implementational complexity, and bitrate are also at the center of our discussions, characterizing many different speech codecs incorporated in source-sensitivity matched wireless transceivers. 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 roadmap for the reader with reference to Cox's work [1, 2]. The formally evaluated mean opinion score (MOS) values of the various codecs described in this book are shown in Figure 1.

Observe in the figure that a range of speech codecs have emerged over the years. These codecs attained the quality of the 64 kbps G.711 PCM speech codec, though at the cost of significantly increased coding delay and implementational complexity. The 8 kbps G.729 codec is the most recent addition to the International Telecommunications Union's (ITU) standard schemes; it significantly outperforms all previous standard ITU codecs in terms of robustness. 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 include 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

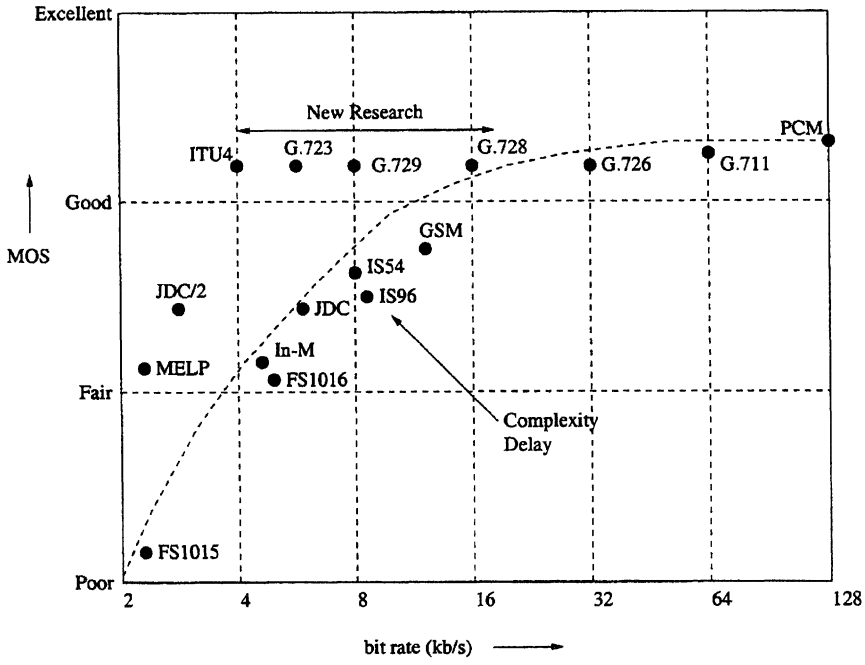

Figure 1 Subjective speech quality of various codecs. Cox *et al.* [1] © IEEE, 1996.

JDC arrangement (JDC/2). These exhibit 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 U.S. Department of Defense 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 hybrid coding family and achieve significant bitrate economies at the cost of increased complexity and delay.

Specifically, the 16 kbps G.728 backward-adaptive scheme maintains a similar speech quality to the 32 and 64 kbps waveform codecs, while also featuring an impressively low, 2 ms delay. This scheme was standardized during the early 1990s. The similar, but significantly more robust, 8 kbps G.729 codec was approved in March 1996 by the ITU. Its standardization overlapped with 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. Work is currently under way to standardize a 4 kbps ITU scheme, which we refer to here as ITU4.

In parallel to the ITU's standardization activities, a range of speech coding standards have been proposed for regional cellular mobile systems. The standardization of the 13 kbps RPE-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. As Figure 0.1 shows, 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 almost coincided with that of the IS-54 arrangement. This codec development was also followed by a half-rate standardization process, leading to the 3.2 kbps Pitch-Synchroneous Innovation CELP (PSI-CELP) scheme.

The 15-95 Pan-American CDMA system also has its own standardized CELP-based speech codec, which is a variable-rate scheme. It supports 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 to be subjectively similar to each other under the perfect channel conditions of Figure 0.1. Finally, 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, though 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 bitrate, quality, and robustness modes, became known in the community. This led to the requirement of designing an appropriate multimode codec, the Advanced Multi- Rate codec referred to as the AMR codec. A range of IMTs are also the subject of this book, as current research on sub-2.4 kbps speech codecs for which auditory masking is more dominant. Lastly, since the wideband codec based on the classic G.722 subband-ADPCM is becoming somewhat obsolete in the light of the exciting new developments in compression, the most recent trend is to consider wideband speech and audio codecs, providing susbtantially enhanced speech quality. As a result of early seminal work on transform-domain or frequency-domain-based compression by Noll and his colleagues, in this field the PictureTel codec (which can be programmed to operate between 10 kbps and 32 kbps and hence is amenable to employment in IMTs), has become the most attractive candidate. The present text portrays this codec in the context of a sophisticated burst-by-burst adaptive wideband turbo-coded Orthogonal Frequency Division Multiplex (OFDM) IMT. This scheme can also transmit high-quality audio signals, behaving essentially as a good waveform codec.

MILESTONES IN SPEECH CODING HISTORY

Over the years a range of excellent monographs and textbooks have been published, characterizing the state-of-the-art and significant milestones. The first major development in the history of speech compression is the invention of the vocoder in 1939. Delta modulation was introduced in 1952 and became well established following Steele's monograph on the topic in 1975 [3]. Pulse Coded Modulation (PCM) was first documented in detail in Cattermole's classic contribution in 1969 [4]. However, in 1967 it was recognized that predictive coding provides advantages over memory-less coding techniques, such as PCM. Predictive techniques were analyzed in depth by Markel and Gray in their 1976 classic treatise [5]. This was followed shortly by the often cited reference [6] by Rabiner and Schafer. In 1979 Lindblom and Ohman also contributed a book on speech communication research [7].

The foundations of auditory theory were laid down as early as 1970 by Tobias [8] but were not fully exploited 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 was comprehensively documented by Jayant and Noll in their 1984 monograph [10]. During the 1980s, speech codec developments were accelerated by the emergence of mobile radio systems, where spectrum was a scarce resource, potentially doubling the number of subscribers and hence the revenues, if the bitrate could be halved.

The RPE principle, as a relatively low-complexity analysis by synthesis technique, was proposed by Kroon, Deprettere, and Sluyter in 1986 [11]. Further research was 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 standardized AbS speech codec, which also employed long-term prediction (LTP), recognizing the important role of pitch determination in efficient speech compression [14, 15]. It was in this period that 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 summarized by, among others, O'Shaughnessy [17], Papamichalis [18], and Deller, Proakis and Hansen [19].

It was also during this era that the importance of speech perception and acoustic phonetics [20] was duly recognized—for example, in the monograph by Lieberman and Blumstein. A range of associated speech quality measures were summarized by Quackenbush, Barnwell, and Clements [21]. Nearly concomitantly, Furui published a book related to speech processing [22]. This period witnessed the appearance of many of the speech codecs seen in Figure 0.1, which found applications in the emerging global mobile radio systems, such as IS-54, JDC. These codecs were typically associated with source-sensitivity matched error protection, for which, for example, Steele, Sundberg, and Wong [23–26] provided early insights on the topic. Further sophisticated solutions were suggested by Hagenauer [27].

During the early 1990s, Atal, Cuperman, and Gersho [28] edited prestigious contributions on speech compression, and Ince [29] contributed a book related to the topic. Anderson and Mohan co-authored a monograph on source and channel coding in 1993 [30]. Most of the recent developments were then consolidated in Kondoz's excellent monograph in 1994 [31] and in the multi-authored contribution edited by Keijn and Paliwal [32] in 1995. The most recent addition to this range of contributions is the second edition of O'Shaughnessy's well-referenced book [19].

PURPOSE AND OUTLINE OF THE BOOK

Against this backdrop (since the publication of Kondoz's monograph in 1994 [31] seven years have elapsed), at the time of writing; this book endeavors to review the recent history of speech compression and communications and to provide the reader with a historical perspective. We begin with a rudimentary introduction to communications aspects, since throughout the book we illustrate the expected performance of the various speech codecs studied in the context of a full wireless transceiver.

The book has four parts. Part I and II cover classic background material, while the bulk of the book comprises research-oriented Parts III and IV, which cover both standardized and proprietary speech codecs and transceivers. Specifically, Part I focuses on classic waveform coding and predictive coding (Chapters 1 and 2). Part II centers on analysis by synthesis-based coding, reviewing the principles in Chapter 3 as well as both narrow and wideband spectral quantization in Chapter 4. RPE and CELP coding are the topic of Chapters 5 and 6, which are followed by a long chapter on the existing forward-adaptive standard CELP codecs in Chapter 7 and on their associated source-sensitivity matched channel coding schemes. Chapter 8 discusses proprietary and standard backward-adaptive CELP codecs, and concludes with a system design example based on a low-delay, multimode wireless transceiver.

The essentially research-oriented Part III is dedicated to a range of standard and proprietary wideband schemes, as well as wireless systems. As an introduction to the scene, the classic G.722 wideband codec is reviewed first, leading to various low-rate wideband codecs. Chapter 9 concludes with a turbo-coded Orthogonal Frequency Division Multiplex (OFDM) wideband audio system design example. The remaining chapters, namely Chapters

10–16 of Part IV, are all dedicated to sub-4 kbps codecs and transceivers. The book is concluded with a brief comparison of a range of various codecs.

This book is limited in terms of its coverage of these aspects simply because of the space limitations. We have nonetheless endeavored to provide the reader with a broad range of applications examples, which are pertinent to a range of typical wireless transmission scenarios.

We hope that the book offers you 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. Analyzing these tradeoffs and proposing a range of attractive solutions to various voice communications problems are basic aims of this book.

Again, it is our hope that the book presents the range of contradictory system design tradeoffs in an unbiased fashion and that you will be able to glean information not only to solve your own particular wireless voice communications problem, but most of all to experience an enjoyable and relatively effortless reading, providing you with intellectual stimulation.

Lajos Hanzo
Clare Somerville
Jason Woodard