Prob. 3

Multistage, or Multiple Stage, Vector Quantization seems to have first appeared in the literature at ICASSP-1982 in the following paper:

Multiple stage vector quantization for speech coding
Biing-Hwang Juang; Gray, A., Jr.; Proc. Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '82. Vol. 7, May 1982, pp. 597–600.

Sample references to MSVQ in standards (and others):

1. MELP
   - Citation:
     Department Of Defense Telecommunications Systems Standard, "Analog-to-digital conversion of voice by 2,400 bit/second mixed excitation linear prediction (MELP)," MIL-STD-3005, 20 Dec 1999.

   - Used for quantization of Line Spectral Frequency (LSF) vectors in low bit-rate real-time speech coding.
   - LSF vectors quantized with multistage vector quantizer in 4 cascaded stages (7-, 6-, 6-, and 6-bit codebooks)
     Number of codewords to compare = 128 + 64 + 64 + 64 = 320 codewords
     For same number of bits, a single stage VQ: 25 bits => 33,554,432 codewords (!)

2. G.722.2
   - Citation:
     ITU-T G.722.2 TELECOMMUNICATION STANDARDIZATION SECTOR OF ITU (01/2002) SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS Digital terminal equipments – Coding of analogue signals by methods other than PCM. Wideband coding of speech at around 16 kbit/s using Adaptive Multi-rate Wideband (AMR-WB)

   - Used for quantization of Immittance Spectral Frequency (ISF) vectors in wideband (7kHz) coding of speech
   - Depending on the bit-rate, combination split-multistage VQ according to the following tables:

BW

Table 2/G.722.2 – Quantization of ISP vector for the 8.85 ,12.65, 14.25, 15.85, 18.25, 19.85, 23.05 or 23.85 kbit/s modes

1. Unquantized 16-element-long ISP vector				
2. Stage 1 (r_1) 8 bits			2. Stage 1 (r_2) 8 bits	
3. Stage 2 $(r^{(2)}_{1,0-2})$ 6 bits	3. Stage 2 $(r^{(2)}_{1,3-5})$ 7 bits	3. Stage 2 $(r^{(2)}_{1,6-8})$ 7 bits	3. Stage 2 $(r^{(2)}_{2,0-2})$ 5 bits	3. Stage 2 $(r^{(2)}_{2,3-6})$ 5 bits

Table 3/G.722.2 – Quantization of ISP vector for the 6.60 kbit/s mode

1. Unquantized 16-element-long ISP vector		
2. Stage 1 (r_1) 8 bits		2. Stage 1 (r_2) 8 bits
3. Stage 2 $(r^{(2)}_{1,0-4})$ 7 bits	3. Stage 2 $(r^{(2)}_{1,5-8})$ 7 bits	3. Stage 2 $(r^{(2)}_{2,0-6})$ 6 bits

3. G.729

- Citation:
  ITU-T G.729 TELECOMMUNICATION STANDARDIZATION SECTOR OF ITU (03/1996) SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)

- Low bit rate, low complexity, toll-quality speech coder using CS-ACELP (Conjugate-Structure Algebraic-Excited Linear-Prediction). Input is band-limited, 8 kHz sampling, 16-bit PCM speech (128 kbs). Used in VoIP, wireless communications, and digital satellite systems.
- The difference between Line Spectral Pairs (LSPs) and their 4th-order MA prediction is quantized using a low complexity two-stage/split VQ: Stage 1, 7-bit codebook; Stage 2, Split 10-bit VQ into two 5-bit codebooks.

4. MPEG-4 (CELP)
   - Citation:
     N1683 MPEG4 Overview
     N1695 MPEG4 Systems FAQ

   - Part of speech coding portion for LSP quantization.
   - A two-stage partial prediction and multistage vector quantization (PPMVQ) is employed for LSP quantization. This quantizer operates either in the standard VQ mode or in the PPM-VQ mode which utilizes interframe prediction, depending on the quantization errors. The standard VQ mode operates as a common two-stage VQ which quantizes the error of the first stage in the second stage.

Not sure of the next three, because I could not access the actual document, but ITU's text search engine returned them on a search for "MSVQ"

5. Audio-visual Objects?
   - Citation:
     ANSI ISO/IEC 14496-3 Information technology Coding of audio-visual objects Part 3: Audio-Adopted by INCITS

6. Digital Radio?
   - Citation:
     CENELEC EN 62272-1
     Digital Radio Mondiale (DRM) Part 1: System specification-IEC 62272-1:2003

7. Digital Radio?
   - Citation:
     DIN DIN EN 62272-1 (DRAFT) Digital Radio Mondiale (DRM) - System specification for digital transmissions in the broadcasting bands below 30 MHz (IEC 103/29/CDV:2002); German version prEN 62272-1:2002, text in English

The following are not standards, per se, but are otherwise interesting applications or variations of MSVQ:

8. Joint, optimal design of MSVQs
   - Citation:
     V. Krishnan, D.V. Anderson, D.V., and K.K. Truong, "Optimal multistage vector quantization of LPC parameters over noisy channels," Speech and Audio Processing, IEEE Transactions on , vol. 12(1), Jan. 2004, pp.1—8.

   - Used for coding of speech in noisy channels.
   - LSFs quantized in 2- and 3- stage codebooks, jointly optimized.

9. Data Hiding
   - Citation:
     Unknown author, Pao-Chi Chang - advisor, Data Hiding Techniques for G.729 and MELP Speech Coding, Date of Defense: June 14, 2002.

   - The author of this Master's thesis used the codebook indices of MSVQ codebooks in MELP and G.729 for **data hiding**, while apparently managing to maintain compatibility with those standards. Unfortunately the paper is in Chinese, so I have not been able to read it.

Abstract
Data hiding is the art of hiding secret messages within a multimedia signal. Most data hiding techniques developed today for speech cannot defense the attack of linear predictive coding (LPC), which is widely used in speech communication systems, and that means it is very difficult to transmit the secret messages and speech simultaneously. A different approach is to hide the secret message in the compressed bit stream. In this thesis, we present two data hiding techniques in compression domain. The first method is called least-significant-bit substitution (LSBS) method. The basic idea of LSBS method is to analyze the significance of each bit of each coded frame, and then substitutes the LSBs with the data to be hidden. The second method is called dither-like data hiding (DDH) method, which utilizes the characteristics of subtractive dithering and the multistage vector quantization (MSVQ) in G.729 and MELP. The secret data is hidden in the index of the MSVQ. The data stream processed by either LSBS method or DDH method is compatible with MELP or G.729 speech coding standard. From the simulation results, both methods can deliver the secret message and the speech signal simultaneously, and the DDH method provides better quality than LSB Substitution Method at the same data embedding rate.

10. Image Coding
    - Citation:
      W. Shigang and C. Hexin, "Multistage vector quantization based on simulated annealing for image coding," IEEE International Conference on Intelligent Processing Systems. ICIPS '97. Vol. 2, Oct. 1997, pp. 1014–1017.

    - Used for faster, and higher SNR in image coding.
    - 2- and 3-stage VQs implemented and modified using simulated annealing process.

11. Coding of Image Sequences
    - Citation:
      B. Hammer, A. Brandt, and M. Schielein, "Hierarchical encoding of image sequences using multistage vector quantization," Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '87, vol. 12, April 1987, pp. 1055–1058.

    - Used for encoding Picture Phone and video conference image sequences
    - 3- and 4-stage Hierarchical MSVQs implemented (scalable coding resulted in remarkable lack of blocking artifacts).

12. Phoneme Recognition and Comparison with Olfaction
    - Citation:
      T. Leen, M. Webb, and S. Rehfuss, "Encoding and classification in a model of olfactory cortex," Neural Networks, 1991, IJCNN-91-Seattle International Joint Conference on, vol. 2, 8–14 July 1991, pp. 553–559.

    - Used for classification of vowels.
    - Neural implementation of 3-stage MSVQ.

16. 16/17

Prob. 4

Scalable coding is the concept or practice of encoding data in such a way as to allow for maximum quality to be obtained under varying conditions, such as individual interface speeds, heavy network traffic, or weather-related packet loss in satellite links. By using such a scheme, a data object need be encoded only once, and then individual users may decode as much as possible, given their particular bandwidth, hardware, or other constraints, and ignore the extra data without catastrophic results.

To achieve this goal, the approach generally taken is to encode the data in different layers: a base layer, which is independent of other layers and must be decoded as it contains the most fundamental data (in some sense); and one or more enhancement layers, which depend on the base layer and may be decoded optionally. This technique would be particularly advantageous in packet-switched networks, where loss or delay of packets might otherwise destroy the quality of streamed multimedia.

Scalability of video is generally categorized into three basic types: Temporal, SNR, and Spatial. Temporal scalability allows video to be viewed at different frame (picture) rates. A common way this is achieved is by encoding the I- and P-frames in the base layer, and encoding the B-frames in an enhancement layer, as they can be safely omitted because the I- and P-frames are not predicted from them.

Video SNR Scalability can be thought of as varying the granularity of a fixed-size picture, whereas Spatial Scalability indicates how well picture quality is maintained when viewed at different sizes. While there is an obvious relationship between the two, they are typically achieved in slightly different ways and are often treated separately as a result. For SNR Scalability, the residual picture error between the encoded image and the original is often encoded and transmitted as an enhancement layer. Spatial Scalability may be accomplished by encoding the base layer so that the decoder can use different interpolation methods for viewing at different resolutions.

Color data in a color video is part of the signal by definition, so variations from (or absence of) that color could be considered noise, so another common implementation of SNR Scalability is encoding the luminance data in the base layer and encoding chrominance data in the enhancement layer (typically YUV format). However, a method soon to be presented at ICASSP next month encodes some color data in all layers via a method called Matching Pursuit, which encodes image data with anisotropic refinement atoms (wavelet-based). Since the method capitalizes on redundancy in the

DSW

data, it works better with RGB format. And although this article is geared toward still images, the concept could certainly be extended to video. Citation:
R. M. Figueras i Ventura, P. Vandergheynst, P. Frossard, and A. Cavallaro, "Color Image Scalable Coding With Matching Pursuit." To be published in IEEE Proceedings of ICASSP 2004.

Yet another type of scalability is Object Scalability, which became possible with the advent of MPEG-4 (and was proposed therein). This technique extends the concept of scalability to the actual content of the data, defining layers based on the priority of each object with respect to their context. For example, in a video of a car race, the objects 'car' and 'track' might be encoded in the base layer, while objects 'cloud' and 'stands' might be included in the enhancement layer. This approach might be even more naturally inclined to MPEG-21.

An idea I have not seen (but it may already exist) would be to include optional "pop-up" text information or graphics in an enhancement layer, similar to object scalability, but where the objects in question are not necessarily part of the picture.

Another idea, which would pit opposing objectives against one another, would be content scalability. I.e. Digital video broadcast viewers would not care if advertisements were mostly encoded in an enhancement layer, because data loss during them is okay. On the other hand, advertisers would probably pay to make sure the situation was reversed. This would obviously depend on forces other than technical feasibility. And it could be related to object scalability as well, where advertisements could be encoded as a different kind of object.

The general concept of scalability could easily be extended to other media types, such as speech or audio coding. Particularly, MSVQ would lend itself naturally to scalable coding, as was observed in the previous problem.