Index

Aboud, G.D., 302
Absolute threshold of hearing (ATH), 165–66
Accelerometers, 30–31
Acoustic event detection (AED), 235, 281–82
Acoustic Long Term Average Spectrum (LTAS), 239
Acoustic pitch, 238
Actuators, 28–29
AdaBoost learning algorithm, 281
Adaptive codebook, 184
Addison, Paul S., 172
ADPCM algorithm, 148–50
ADRS envelope, 47
Advanced perceptual compression convolution theorem, 160–61
Cooley-Tukey algorithm, 157
Discrete Cosine Transform (DCT), 158–60
Discrete Fourier Transform (DFT), 156–58
Fast Cosine Transform, 157
Fast Fourier Transform, 157
JPEG format and, 161–64
overview, 156
perceptual video compression, 170–71.
(see also Perceptual video compression)
psychoacoustics and, 164–68
AED (Acoustic event detection), 235, 281–82
Ahmed, N., 171–72
A-law, 145–46
Algorithms
AdaBoost learning algorithm, 281
ADPCM algorithm, 148–50
arithmetic coding, 135–38, 136t.11.5
Cooley-Tukey algorithm, 157
Huffman coding, 128–31
k-means algorithm, 142–43
Lempel-Ziv algorithms, 131–35
LZ77 algorithm, 131–33, 132t.11.3
LZ78 algorithm, 133–36, 133t.11.4
LZC algorithm, 135
LZW algorithm, 135
machine learning algorithms, 235–36
machine vision algorithms, 235
MP3 format, 168–70
MPEG-1 algorithm, 151–53
Sequential Minimal Optimization (SMO) algorithm, 257
Viterbi algorithm, 259, 272–73
x-means algorithm, 143–44
Allen, J.F., 34
American National Standards Institute (ANSI), 53
Anderson, C.A., 121
ANNs (Artificial Neural Networks), 252–55
Anstis, S.M., 302
Apple, 108
Apple iTunes, 202
Application layer QoS control, 100
Approximants, 44
Arithmetic coding, 135–38, 136t.11.5
Artificial intelligence, 236
Artificial Neural Networks (ANNs), 252–55
ATH (Absolute threshold of hearing), 165–66
Attack, 47
Audio editing, 73–74
Audio processing, 235
Audio sensors, 29
Audiovisual media, 7–8
Autostereoptic displays, 59
A-weighting scheme, 37
Baars, B.J., 93
Background, signal processing and, 223–24
Bark scale, 165, 217
Barkhausen, Heinrich, 165
Barrow, H.G., 302
Bayesian Information Criterion (BIC), 252, 262, 279–80
Belongie, S., 302
Berkeley, George, 290, 302
Bickford, A.C., 49
Big Data, 11
Binary Format for Scenes (BIFS), 171
Bing, 297
Blackstock, D.T., 48
Blue-ray format
  compression and, 124
differential coding and, 151
  lossy compression and, 141
  Body transfer illusion, 84
Bohr, Niels, 305
Books, 67–68
Botvinick, M., 84
Bovik, Al, 231
Brandenburg, K., 172
Broughton, S.A., 171–72
Bryan, K., 171–72
Bunke, Horst, 287
Bureau, M., 121
Bushman, B.J., 121
Caitlin, D.E., 25
Calvert, Gemma, 93
Cameras, 54–57
  autostereoptic displays, 59
digital cameras, 57
  Exif format and, 24–25
  image formation, 54–56
  line of sight, 56
  moving camera, moving objects (MCMO), 282–83
  moving camera, stationary objects (MCSO), 282–83
  parallax barrier method, 59
  resolution, 56
  stationary camera, moving objects (SCMO), 282–83
  stationary camera, stationary objects (SCSO), 282–83
television and, 56–57
  3D video, 57, 59
video cameras, 56–57
Candelas, 52
Carrier, J. Edward, 34
CBIR. See Content based image retrieval (CBIR)
CELP. See Code Excited Linear Prediction (CELP)
Challenges in multimedia systems
  context versus content, 23–25
  semantic gap, 22–23, 25
  overview, 21
Chroma quantization, 162
CIELAB color space, 62–64
CIEXYZ color space, 63–64
Clark, A.B., 146
Classic information retrieval
  logical representation, 193–94
Multimedia Information Retrieval (MIR)
  compared, 192–98
  overview, 192–94
PageRank, 194–95
  precision, 194
  ranking, 194
recall, 194
  relevance feedback, 194, 196–98
term frequency/inverse document frequency
  (TF/IDF), 195–96
Web Image Search, 198
Cloud computing, 11, 308
CMYK color space, 61
Codd, Edgar, 189–90
Code Excited Linear Prediction (CELP)
  adaptive codebook, 184
  bitstream, 185t.14.1
  fixed codebook, 184–85
  line spectral frequencies (LSF), 183–84
  line spectral pairs (LSP), 183–84
  overview, 182–83
Cohen, J.D., 84
Color histograms, 243–44
Color spaces, 61–64
  CIELAB color space, 62–64
  CIEXYZ color space, 63–64
  CMYK color space, 61
  RGB color space, 61
  YUV color space, 62
Color vision, 60–61
Communication
defined, 16
evolution of technology, 8–10
  8t.2.1
in human society, 7–8
  overview, 25
Compasses, 30–31
Compression
  advanced perceptual compression. (see Advanced perceptual compression)
  algorithms, 128–38
  arithmetic coding, 135–38, 136t.11.5
  DEFLATE method, 133
  dynamic range compression (DRC), 216–17
  entropy and, 127
  Huffman coding, 128–31
  information content and, 125–32, 126t.11.1,
  127t.11.2
  Lempel-Ziv algorithms, 131–35
  Lossy compression. (see Lossy compression)
  LZ77 algorithm, 131–33, 132t.11.3
  LZ78 algorithm, 133–36, 133t.11.4
  LZW algorithm, 135
  overview, 4, 124
  perceptual video compression, 170–71
  (see also Perceptual video compression)
  run-length encoding (RLE), 124–25
  source coding theorem, 127
  speech compression. (see Speech compression)
  weakness of entropy-based compression
  methods, 138
  Computer graphics, 235

© in this web service Cambridge University Press
www.cambridge.org
### Index

**Computing**
- evolution of technology, 8–10, 9t.2.2
- expected results of computing, 11–12
- personal computers, 10–11

**Connected components, signal processing and**, 223

**Content analysis. See Multimedia content analysis**

**Content analysis systems. See Multimedia content analysis systems**

**Content based image retrieval (CBIR),** 199–205
- overview, 199–201
- Query by Humming (QbH), 202
- semantic content-based retrieval, 202–3
- tag-based retrieval, 203–4
- video retrieval, 204–5

**Content replication,** 101

**Content segments,** 69

**Context versus content**
- connecting data and users, 292
- context defined, 293–94
- context in content, 294, 295–96
- context only image search, 297
- data acquisition context, 294–95
- device parameters, 297–99
- interpretation context, 295
- overview, 23–25, 290–91
- perceivers, 299–302
- smart phone photo management, 300–2
- types of context, 294–95
- verification vision, 296–97

**Continuous media distribution services, 100–1**

**Contours,** 227

**Control theory,** 25

**Conventional video encoding, 170–71**

**Convolution theorem, 160–61**

**Cooley-Tukey algorithm, 157**

**Countermeasures regarding privacy, 118**

**Crocker, Lee Daniel, 151**

**CRT technology, 58**

**Data acquisition and organization in documents, 71–72**

**Data acquisition context, 294–95**

**Databases**
- logical level of data, 190
- Multimedia Information Retrieval (MIR) and, 189–91
- organization, storage, management, and retrieval (OSMR), 189–90
- physical level of data, 190
- view level of data, 190–91

**Daubechies, Ingrid, 172**

**DCT (Discrete Cosine Transform), 158–60**

**Decay, 47**

**Decibel scale, 37**

**Defining multimedia systems**
- audiovisual media and, 7–8
- communication in human society, 7–8
- elements of multimedia computing, 7–8
- emerging applications and, 11
- evolution of computing and communication technology, 8–10, 9t.2.1, 9t.2.2
- evolving nature of information and, 11–12
- expected results of computing and, 11–12
- experiential environments and, 12–13, 14
- formal definition, 13
- multimedia aspect, 10–13
- personal computers and, 10–11
- printing press and, 7, 13
- storage and recording technology and, 8
- telephony and, 7
- telephony and, 10

**DEFLATE method, 133**

**DER (Diarization Error Rate), 280**

**Desktop metaphor, 111, 112**

**Detection error tradeoff (DET) curve, 264**

**Device parameters, 297–99**
- derived metal layer, 298
- human induced metal layer, 298
- meta layer, 297
- optical meta layer, 297
- overview, 294
- pixel/spectral layer, 297
- spatial metal layer, 298
- temporal metal layer, 297

**DFT (Discrete Fourier Transform), 156–58**

**Dialog boxes, 110**

**Diarization Error Rate (DER), 280**

**Dictionaries, 16–17**

**Differential coding, 147–53**
- ADPCM algorithm, 148–50
  - in audio, 148–50
  - in images, 150–51
- MPEG-1 algorithm, 151–53
- Paeth filter, 150–51
- PNG format, 150–51, 151t.12.1
  - in video, 151–53

**Digital cameras**
- Exif format and, 24–25, 161, 298–99
- overview, 57

**Digital television**
- differential coding and, 151
- lossy compression and, 141

**Digital video discs (DVDs)**
- differential coding and, 151
- lossy compression and, 141

**Digitization, 32–33**
- discretization error, 33
- Nyquist frequency, 33
- quantization, 32
- quantization noise, 33
- sampling, 32
<table>
<thead>
<tr>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diphthongs, 44</td>
</tr>
<tr>
<td>Discrete Cosine Transform (DCT), 158–60</td>
</tr>
<tr>
<td>Discrete Fourier Transform (DFT), 156–58</td>
</tr>
<tr>
<td>Discretization error, 33</td>
</tr>
<tr>
<td>Distraction, 120</td>
</tr>
<tr>
<td>Dix, Alan, 114, 121</td>
</tr>
<tr>
<td>Documents</td>
</tr>
<tr>
<td>audio editing, 73–74</td>
</tr>
<tr>
<td>audiovisual technology and, 68–69</td>
</tr>
<tr>
<td>books, 67–68</td>
</tr>
<tr>
<td>content segments, 69</td>
</tr>
<tr>
<td>current authoring environments, 78–79</td>
</tr>
<tr>
<td>data acquisition and organization, 71–72</td>
</tr>
<tr>
<td>defined, 67–69</td>
</tr>
<tr>
<td>dynamic documents, 71</td>
</tr>
<tr>
<td>editing, 72</td>
</tr>
<tr>
<td>elements of multimedia authoring environment, 75–77</td>
</tr>
<tr>
<td>emerging multimedia editing tools, 74</td>
</tr>
<tr>
<td>evolving nature of, 69–71</td>
</tr>
<tr>
<td>HTML and, 78–79</td>
</tr>
<tr>
<td>linear nature of, 69</td>
</tr>
<tr>
<td>media asset characteristics, 75</td>
</tr>
<tr>
<td>MPEG4 and, 79</td>
</tr>
<tr>
<td>non-linear flow, 71</td>
</tr>
<tr>
<td>overview, 67</td>
</tr>
<tr>
<td>photo editing, 73, 73n.2</td>
</tr>
<tr>
<td>Photoshop and, 79</td>
</tr>
<tr>
<td>publishing formats, 77</td>
</tr>
<tr>
<td>representations of multimedia documents, 77</td>
</tr>
<tr>
<td>selection, 72</td>
</tr>
<tr>
<td>SMIL and, 79</td>
</tr>
<tr>
<td>spatial layout, 75–76</td>
</tr>
<tr>
<td>stages in document creation, 71–74</td>
</tr>
<tr>
<td>structure-based representation, 77</td>
</tr>
<tr>
<td>synchronization, 76–77</td>
</tr>
<tr>
<td>temporal layout, 76</td>
</tr>
<tr>
<td>text editing, 72–73</td>
</tr>
<tr>
<td>time-based representation, 77</td>
</tr>
<tr>
<td>type of media and, 69–71</td>
</tr>
<tr>
<td>video editing, 74</td>
</tr>
<tr>
<td>Dodd, George, 34</td>
</tr>
<tr>
<td>Dolby AAC, 165</td>
</tr>
<tr>
<td>Download and play, 100</td>
</tr>
<tr>
<td>Downsampling, 217–18</td>
</tr>
<tr>
<td>DRC (Dynamic range compression), 216–17</td>
</tr>
<tr>
<td>Duda, Richard O., 269</td>
</tr>
<tr>
<td>Durbin, J., 180</td>
</tr>
<tr>
<td>Durham, M., 121</td>
</tr>
<tr>
<td>DVDs (Digital video discs)</td>
</tr>
<tr>
<td>differential coding and, 151</td>
</tr>
<tr>
<td>lossy compression and, 141</td>
</tr>
<tr>
<td>Dynamic documents, 71</td>
</tr>
<tr>
<td>Dynamic Huffman compressors, 131</td>
</tr>
<tr>
<td>Dynamic range compression (DRC), 216–17</td>
</tr>
<tr>
<td>E-Chalk lectures, 85</td>
</tr>
<tr>
<td>Echo, 38</td>
</tr>
<tr>
<td>Edges, 226–28</td>
</tr>
<tr>
<td>contours, 227</td>
</tr>
<tr>
<td>detection of, 228</td>
</tr>
<tr>
<td>edge detectors, 227</td>
</tr>
<tr>
<td>edge fragments, 227</td>
</tr>
<tr>
<td>edge linking, 227</td>
</tr>
<tr>
<td>edge orientations, 227</td>
</tr>
<tr>
<td>edge points, 227</td>
</tr>
<tr>
<td>enhancement of, 227</td>
</tr>
<tr>
<td>filtering of, 227</td>
</tr>
<tr>
<td>isotropic operators, 227</td>
</tr>
<tr>
<td>line discontinuities, 226</td>
</tr>
<tr>
<td>ramp edges, 226</td>
</tr>
<tr>
<td>roof edges, 226</td>
</tr>
<tr>
<td>step discontinuities, 226</td>
</tr>
<tr>
<td>vectors, 227</td>
</tr>
<tr>
<td>Edison, Thomas A., 39</td>
</tr>
<tr>
<td>Edit distance, 264–66</td>
</tr>
<tr>
<td>Editing of documents, 72</td>
</tr>
<tr>
<td>Einstein, Albert, 305</td>
</tr>
<tr>
<td>Elements of multimedia computing events, 17–18, 25–26</td>
</tr>
<tr>
<td>experience, 15–17</td>
</tr>
<tr>
<td>information, 15–17</td>
</tr>
<tr>
<td>objects, 17–18, 25–26</td>
</tr>
<tr>
<td>overview, 15</td>
</tr>
<tr>
<td>perception, 18–19</td>
</tr>
<tr>
<td>perceptual cycle, 19–21</td>
</tr>
<tr>
<td>EM (Expectation Maximization), 251</td>
</tr>
<tr>
<td>Eng, Per, 34</td>
</tr>
<tr>
<td>Entropy</td>
</tr>
<tr>
<td>compression and, 127</td>
</tr>
<tr>
<td>multimedia content analysis and, 262</td>
</tr>
<tr>
<td>Environmental sensors, 30–31</td>
</tr>
<tr>
<td>Estimation theory, 25</td>
</tr>
<tr>
<td>Euclidean distance, 261–62</td>
</tr>
<tr>
<td>Events</td>
</tr>
<tr>
<td>as element of multimedia computing, 17–18, 25–26</td>
</tr>
<tr>
<td>event oriented thinking, 18</td>
</tr>
<tr>
<td>Everst, F. Alton, 48</td>
</tr>
<tr>
<td>Exchangeable Image File (Exif) format</td>
</tr>
<tr>
<td>cameras and, 24–25, 161, 298–99</td>
</tr>
<tr>
<td>overview, 26, 80</td>
</tr>
<tr>
<td>privacy and, 117</td>
</tr>
<tr>
<td>smart phones and, 302</td>
</tr>
<tr>
<td>tag-based retrieval and, 203–4</td>
</tr>
<tr>
<td>Exotic sensors, 314</td>
</tr>
<tr>
<td>Expectation Maximization (EM), 251</td>
</tr>
<tr>
<td>Experience</td>
</tr>
<tr>
<td>defined, 16</td>
</tr>
<tr>
<td>as element of multimedia computing, 15–17</td>
</tr>
<tr>
<td>future trends, 310–12</td>
</tr>
<tr>
<td>Experiential environments and, 12–13, 14</td>
</tr>
<tr>
<td>Explicit specification of synchronization, 90–91</td>
</tr>
<tr>
<td>Extensible Markup Language (XML), 24, 191–92</td>
</tr>
<tr>
<td>Extensible MPEG-4 Textual (XMT), 171</td>
</tr>
</tbody>
</table>
## Index

Face detection, 280–81
Face recognition, 117, 280–81
Facebook, 314
False creativity, 111–12
Fast Cosine Transform, 157
Fast Fourier Transform, 157
Ferreira, A.J., 172
Figure-ground problem, 212
Filter by example, 220–22
Fixed codebook, 184–85
Flash, 118–19
Flickr.com, 234
Floyd, R., 49
Fonts, 110
Formants, 45, 239
Fostering learning, 112–14
Fourier, Jean-Baptiste-Joseph, 156
Fraden, Jacob, 34
Frames, 209
Frank, E., 269
Fredric, Harris, 172
Frequency/loudness resolution, 165–66
Frequency space quantization, 162–63
Freuder, E.C., 302
Fricative consonants, 44
Friedland, Gerald, 121, 303
F-score, 264
Fundamental frequency, 47
Fundamental properties of multimedia applications and systems
documents. (see Documents)
multimedia systems. (see Multimedia systems)
multimodal integration. (see Multimodal integration)
overview, 4
privacy, 116–18
safety, 119–20
security, 118–19
synchronization. (see Synchronization)
user interface design principles. (see User interface design principles)
Future trends
experience, 310–12
mobility, 309–10
networking, 308
overview, 4, 305, 315
prediction methods, 305–6
privacy, 314–15
semantic computing, 312–14
sensors, 308–9
transistors, 307–8
Galleguillos, C., 302
Gauntlett, David, 121
Gaussian Mixture Models (GMMs), 250–52, 272–73, 274, 277, 279–80, 281–82
Gaussianization, 273–74
German Telefunken, 58
GIF (Graphics Interchange Format), 135
GIST features, 246
Gleick, James, 5, 14, 25, 79
Global level of computation, 211
Global positioning systems (GPS), 30–31
Global System for Mobile (GSM), 185–86
GMMs. See Gaussian Mixture Models (GMMs)
Gold, R., 287
Gold, Ben, 48–49
Gold, Bernard, 231
Google, 297
GPS (Global positioning systems), 30–31
Graphical filters, 222
Graphical user interface (GUI), 108, 111–12
Graphics Interchange Format (GIF), 135
Gregory, Richard, 302
GSM (Global System for Mobile), 185–86
GUI (Graphical user interface), 108, 111–12
Guild, J., 64
Gutenberg, Johannes, 7, 13, 67, 79
Haar Wavelet basis functions, 281
Haptic technology, 28–29
Harmonicity-to-Noise Ratio (HNR), 238
Harold, Elliotte Rusty, 121
Hart, Peter E., 269
Hayakawa, Samuel Ichiye, 5
HCI (Human-computer interaction), 108–9
HDTV, 59, 124
Headset microphones, 39–40
Hershenson, M., 93
Hershenson experiments, 84
Hidden Markov Model (HMM), 257–58, 272–73, 274, 275–76, 281–82
Hipparchus, 146
HNR (Harmonicity-to-Noise Ratio), 238
Holes, signal processing and, 223–24
Holographic light sources, 64
HTML (HyperText Markup Language), 78–79
Huffman, David A., 128, 138
Huffman coding, 128–31
Huffman tree, 128
Human sensors, 28–29
Human-computer interaction (HCI), 108–9
Hype Cycle, 305–6
HyperText Markup Language (HTML), 78–79
IEC (International Electrochemical Commission), 37
IMA (Interactive Multimedia Association), 148
Image processing, 235
Immersive video, 105, 106
Implicit specification of synchronization, 90
Indexing, 199
Information
defined, 16
as element of multimedia computing, 15–17
## Index

Information content and compression, 125–32, 126.11.1., 127.11.2

Information retrieval
classic information retrieval. (see Classic information retrieval)
Multimedia Information Retrieval. (see Multimedia Information Retrieval (MIR))
Input devices, 97–98
Integrative systems approach, 1–3
Interaction environment, 199
Interactive Multimedia Association (IMA), 148
International Electrochemical Commission (IEC), 37
Interpolation
linear interpolation, 218
polynomial interpolation, 219
spline interpolation, 219
Interpolation filters, 218–19
Interpretation context, 295
Intra-object specification of synchronization, 91
iPhoto, 203–4
IP Multicast, 101
Isotropic operators, 227
Jain, Ramesh, 14, 25–26, 79, 231
Jeoung, Ye Sun, 302
JFIF (JPEG File Interchange Format), 161
Johnson, Jeff, 120
Johnston, J.D., 172
Joint Photographic Experts Group (JPEG)
format
advanced perceptual compression and, 161–64
arithmetic coding and, 135
chroma quantization, 162
frequency space quantization, 162–63
linearization, 163–64
lossless encoding, 163–64
tiling, 162
visual quantization and, 147
Jones, Gerard, 121
Jones, Matt, 121
Jones, M.J., 281
JPEG File Interchange Format (JFIF), 161
JPEG format. See Joint Photographic Experts Group (JPEG) format
Juang, B.H., 231
Kan, Min Yen, 93
Kellner, D., 121
k-Gaussians, 249–50
Kientzle, Tim, 48
Kittler, J., 302
KL-Divergence, 262
k-means, 249–50
k-means algorithm, 142–43
k-nearest neighbors, 248–49
Knoll, Thomas, 79
Knuth, Donald, 269
Kullback-Leibler distance, 282
LAB (Look-ahead buffer), 131
LAME, 168–70
Large vocabulary automatic speech recognition (LVASR), 273–75
decoder, 274
feature extraction, 273
feature normalization, 273–74
recognition, 274
speech activity detection, 273
textual postprocessing, 274–75
Lavalier microphones, 39–40
Layout frames, 92
Lempel, Abraham, 131, 138
Lempel-Ziv algorithms, 131–35
Lens distortion, 54
Levinson, N., 180
Levinson-Durbin recursion, 181
Life sensors, 30
Light
additive methods, 58
Cameras, 54–57. (see also Cameras)
candelas, 52
color spaces, 61–64
CRT technology, 58
damage caused by, 120
defined, 51–53
HDTV, 59
holographic light sources, 64
diffusion, 54
lumens, 53
luxes, 53
measurement of, 52–53
optics, 54
overview, 51
perception of, 59–61
polarization, 52
production of, 64
properties of, 53–54
recording of, 54–57
reflection, 53
refraction, 53
reproduction of, 58–59
subtractive methods, 58
television. (see Television)
wave nature of, 51–52
Line discontinuities, 226
Line spectral frequencies (LSF), 183–84
Line spectral pairs (LSP), 183–84
Linear filters, 215–16
Linear interpolation, 218
Linear nature of documents, 69
Linear operators, 214–16
Linear Prediction Coefficients (LPCs), 239
Linear Protective Coding (LPC), 175–82
Index

analysis, 179–82
Levinson-Durbin recursion, 181
Magnitude-Difference Function, 178
Magnitude-Sum Function, 176–77
overview, 175–76
pitch estimation, 177–79
Schuer recursion, 181
voiced/unvoiced detection, 176–77
Zero-Crossing-Rate, 177
Line quantization, 142
Linearization, 163–64
Linux, 146
Live video
immersive video, 105, 106
multiple camera videos produced as single, 103
Multiple Perspectives Interactive (MPI) video, 104–5
one camera video, 103
overview, 103
Lo, W.T., 121
Local level of computation, 210–11
Logical representation, 193–94
Lombard effect, 44
Look-ahead buffer (LAB), 131
Lossless encoding, 163–64
Lossy compression
A-law, 145–46
differential coding, 147–53. (see also Differential coding)
k-means algorithm, 142–43
linear quantization, 142
motion quantization, 147
μ-law, 145–46
overview, 141
perceptual quantization, 144–47
sound amplitude quantization, 145–46
vector quantization, 141–44
visual quantization, 146–47
x-means algorithm, 143–44
Low, Andrew, 245
LPC. See Linear Protective Coding (LPC)
LPCs (Linear Prediction Coefficients), 239
LSF (Line spectral frequencies), 183–84
LSP (Line spectral pairs), 183–84
LTAS (Acoustic Long-Term Average Spectrum), 239
Lumens, 53
Luxes, 53
LVASR. See Large vocabulary automatic speech recognition (LVASR)
LZ77 algorithm, 131–33, 132t.11.3
LZ78 algorithm, 133–36, 133t.11.4
LZC algorithm, 135
LZW algorithm, 135
MacDonald, J., 83
Machine learning algorithms, 235–36
Machine vision algorithms, 235
Magnitude-Difference Function, 178
Magnitude-Sum Function, 176–77
Malware, 118–19
Manhattan distance, 261–62
Marsden, Gary, 121
Martinville, Édouard-Léon Scott de, 39
Masking, 166–68
Maximum Likelihood Linear Regression (MLLR), 274
McGurk, H., 83, 93
McGurk effect, 83
McKay, David, 138
McLuhan, Marshall, 13–14
Media asset characteristics, 75
Media processing to extract features, 199
Media synchronization, 102
Mel Frequency Cepstral Coefficients (MFCCs), 239–42, 272–73, 279
Mel scale, 165, 217
Menu bars, 108
Metadata, 23–25, 117
Microphones, 39–41
headset microphones, 39–40
lavalier microphones, 39–40
noise-canceling microphones, 40
omnidirectional microphones, 40–41
parabolic microphones, 40
as sensors, 29
Microsoft, 108
Microsoft WMA, 165
MIDI, 48, 74
Minkowski metric, 261–62
MIR. See Multimedia Information Retrieval (MIR)
Misra, Pratap, 34
MLLR (Maximum Likelihood Linear Regression), 274
MLPs (Multi-Layer Perceptrons), 252–55
Mobility, future trends, 309–10
Modern video encoding, 171
Monomedia content analysis, evolution to multimedia content analysis, 234–36
Monophthongs, 44
Moore, Andrew, 144
Moore’s Law, 307–8
Motion pictures, 7–8
Motion pictures, 7–8
Motion pictures, 7–8
Motion quantization, 147
MP3 format
advanced perceptual compression and, 168–70
algorithm, 168–70
LAME, 168–70
variable gain amplifiers and, 217
MPEG-1, 151–53, 165, 171, 307
MPEG-2, 165, 171
MPEG-4, 79, 171, 286
MPEG-7, 171
MPEG-21, 171
MPI (Multiple Perspectives Interactive) video, 104–5
Multi-Layer Perceptrons (MLPs), 252–55
Multimedia content analysis
accuracy, 262–63
acoustic event detection, 235
Acoustic Long-Term Average Spectrum (LTAS), 239
acoustic pitch, 238
artificial intelligence, 236
Artificial Neural Networks (ANNs), 252–55
bias, 267
color histograms, 243–44
computer graphics, 235
correlation versus causation, 266
data collection and annotation, 267
domain independence, 267–68
edit distance, 264–66
energy/intensity, 238
entropy and, 262
error measurement and evaluation, 262–66
Expectation Maximization (EM), 251
features of, 238–47
formants, 239
F-score, 264
Gaussian Mixture Models, 250–52
GIST features, 246
Harmonicity-to-Noise Ratio (HNR), 238
Hidden Markov Model (HMM), 257–58
k-Gaussians, 249–50
k-nearest neighbors, 248–49
Linear Prediction Coefficients (LPCs), 239
machine learning algorithms, 235–36
machine vision algorithms, 235
Mel Frequency Cepstral Coefficients (MFCCs), 239–42
monomedia content analysis, evolution from, 234–36
Multi-Layer Perceptrons (MLPs), 252–55
music analysis, 235
music synthesis, 235
optical flow, 246–47
overtraining, 268
overview, 233
pattern recognition, 235–36
Perceptrons, 252–55
precision/recall, 263–64
psychophysics, 236
ROC/DET curve, 264
Scale-Invariant Feature Transformation (SIFT), 245–46
semantic computing, 236
Sequential Minimal Optimization (SMO) algorithm, 257
setup of experiment, 236–38
significance of results, 268
soft margin classification, 257
sparsity, 267
speech analysis, 235
speech synthesis, 235
supervised learning, 248–59
Support Vector Machines (SVMs), 255–57
systems. (see Multimedia content analysis systems)
temporal modeling, 257–59
undermodeling, 268
unsupervised modeling, 260–62
visual texture, 242–43. (see also Visual texture)
Viterbi algorithm, 259
Word Error Rate (WER), 264–66
Multimedia content analysis systems
acoustic event detection (AED), 281–82
face detection, 280–81
face recognition, 280–81
large vocabulary automatic speech recognition (LVASR), 273–75. (see also Large vocabulary automatic speech recognition (LVASR))
optional character recognition (OCR), 275–76. (see also Optional character recognition (OCR))
overview, 272, 286–87
speaker diarization, 278–80. (see also Speaker diarization)
speaker recognition, 276–78. (see also Speaker recognition)
speech activity detection, 272–73
visual object recognition, 282–86
Multimedia Information Retrieval (MIR)
classic information retrieval compared, 192–98
cost based image retrieval (CBIR), 199–205
databases and, 189–91
indexing, 199
information defined, 189
interaction environment, 199
media processing to extract features, 199
multimedia defined, 189
overview, 188–89, 205–6
Query by Humming (QbH), 202
retrieval defined, 199
semantic content-based retrieval, 202–3
structured data, 191–92
tag-based retrieval, 203–4
unstructured data, 191–92
video retrieval, 204–5
Multimedia systems
components of, 95–98
configurations of multimedia nodes, 99
content analysis systems. (see Multimedia content analysis systems)
defining. (see Defining multimedia systems)
download and play, 100
emerging systems, 105
immersive video, 105, 106
input devices, 97–98
live video. (see Live video)
multiple camera videos produced as single, 103
Multiple Perspectives Interactive (MPI) video, 104–5
networking, 98
one camera video, 103
output devices, 97–98
overview, 95, 105–6
processing unit, 96
Quality of Experience (QoE), 99
Quality of Service (QoS), 99
storage, 97
stored video, 100
streaming multimedia. (see Streaming multimedia)

Multimodal integration
body transfer illusion, 84
McGurk effect, 83
overview, 95, 105–6
sensor integration, 86
split attention, 84–86
uncertainty reduction, 84
ventriloquism, 83–84

Multiple camera videos produced as single, 103
Multiple Perspectives Interactive (MPI) video, 104–5
Music, 46–48
ADRS envelope, 47
attack, 47
decay, 47
fundamental frequency, 47
Music Genome Project, 203
Pandora, 203
percussion instruments, 48
Query by Humming (QbH), 202
release, 47
spectrum, 47
string instruments, 47
sustain, 47
timbre, 47
wind instruments, 47
Music analysis, 235
Music synthesis, 235
Mussen, P., 121
μ-law, 145–46
MYSYS system, 295–96

Nahrstedt, Klara, 93, 106
National Institute of Standards and Technology (NIST), 280
Neighbors, signal processing and, 222–23
Neisser, Ulrich, 20, 25, 302–3
Network filtering, 101
Networking
future trends, 308
overview, 98
Nielsen, Jakob, 120–21
Nintendo, 59
Noise-canceling microphones, 40
Non-linear filters, 216–19
Non-linear flow of documents, 71
Norman, Donald A., 120
Norvig, Peter, 269
Nyquist, Harry, 33n.1, 34
Nyquist frequency, 33, 41
Nyquist-Shannon sampling theorem, 33n.1
Object-background separation, 213
Object level of computation, 211
Objects
as element of multimedia computing, 17–18, 25–26
intra-object specification, 91
object oriented thinking, 17
OCR. See Optional character recognition (OCR)
Offset, 131
Ogg Vorbis, 165
Oliva, Aude, 246
Omnidirectional microphones, 40–41
One camera video, 103
Oppenheim, Alan V., 171–72, 231
Optical flow, 246–47
Optics, 54
Optional character recognition (OCR), 275–76
binarization, 275
decoder, 275–76
recognition, 275
rectification, 275
Organization and analysis of multimedia content
multimedia content analysis. (see Multimedia content analysis)
multimedia content analysis systems,
(see Multimedia content analysis systems)
Multimedia Information Retrieval (MIR),
(see Multimedia Information Retrieval (MIR))
overview, 4
signal processing. (see Signal processing)
Output devices, 97–98
Overdriven signals, 41

Paeth filter, 150–51
PageRank, 194–95
Pandora, 203
Parabolic microphones, 40
Parallax barrier method, 59
Path, Signal processing and, 223
Pattern recognition, 235–36
Pelleg, Dan, 144
Perceivers, 299–302
domain knowledge, 300
overview, 295
smart phone photo management, 300–2
Perception
defined, 18
as element of multimedia computing, 18–19
models and, 19
perceptual cycle, 19–21
Perceptrons, 252–55
Perceptual cycle, 19–21
### Index

<table>
<thead>
<tr>
<th>Perceptually encoded information</th>
<th>Quality of Experience (QoE), 99</th>
</tr>
</thead>
<tbody>
<tr>
<td>light. (see Light)</td>
<td>Quality of Service (QoS), 99</td>
</tr>
<tr>
<td>overview, 3</td>
<td>Quantization</td>
</tr>
<tr>
<td>sensors. (see Sensors)</td>
<td>chroma quantization, 162</td>
</tr>
<tr>
<td>sound. (see Sound)</td>
<td>digitization and, 32</td>
</tr>
<tr>
<td>Perceptual quantization, 144–47</td>
<td>frequency space quantization, 162–63</td>
</tr>
<tr>
<td>A-law, 145–46</td>
<td>linear quantization, 142</td>
</tr>
<tr>
<td>motion quantization, 147</td>
<td>motion quantization, 147</td>
</tr>
<tr>
<td>μ-law, 145–46</td>
<td>perceptual quantization, 144–47</td>
</tr>
<tr>
<td>sound amplitude quantization, 145–46</td>
<td>(see also Perceptual quantization)</td>
</tr>
<tr>
<td>visual quantization, 146–47</td>
<td>signal processing and, 209–10</td>
</tr>
<tr>
<td>Perceptual video compression, 170–71</td>
<td>sound amplitude quantization, 145–46</td>
</tr>
<tr>
<td>conventional video encoding, 170–71</td>
<td>vector quantization, 141–44. (see also Vector quantization)</td>
</tr>
<tr>
<td>modern video encoding, 171</td>
<td>visual quantization, 146–47</td>
</tr>
<tr>
<td>Percussion instruments, 48</td>
<td>Quantization noise, 33</td>
</tr>
<tr>
<td>Peregrin, Jaroslav, 25</td>
<td>Query by Humming (QbH), 202</td>
</tr>
<tr>
<td>Persaud, Krishna, 34</td>
<td>Query by Image Example, 200</td>
</tr>
<tr>
<td>Person detection, 314</td>
<td>Questionnaires, 115</td>
</tr>
<tr>
<td>Personal computers, 10–11</td>
<td>Quinton, A., 25</td>
</tr>
<tr>
<td>Phonautograms, 39</td>
<td>Rabiner, Lawrence R., 231</td>
</tr>
<tr>
<td>Phonemes, 44</td>
<td>Radio, 7–8</td>
</tr>
<tr>
<td>Phonograph cylinders, 39</td>
<td>Ramachandran, V.S., 302</td>
</tr>
<tr>
<td>Photo editing, 73, 73n.2</td>
<td>Ramp edges, 226</td>
</tr>
<tr>
<td>Photoshop, 79</td>
<td>Ranking in classic information retrieval, 194</td>
</tr>
<tr>
<td>Pitch estimation, 177–79</td>
<td>Raskin, Jef, 120–21</td>
</tr>
<tr>
<td>Pixels, 209</td>
<td>Real Time Control Protocol (RTCP), 102–3</td>
</tr>
<tr>
<td>Platt, John C., 257</td>
<td>Real Time Streaming Protocol (RTSP), 103</td>
</tr>
<tr>
<td>PNG. See Portable Network Graphics (PNG)</td>
<td>Real Time Transport Protocol (RTP), 102–3, 171</td>
</tr>
<tr>
<td>Pohlmann, Ken C., 48, 153</td>
<td>Recall in classic information retrieval, 194</td>
</tr>
<tr>
<td>Polarization, 52</td>
<td>Receiver-operator-characteristics (ROC) curve, 264</td>
</tr>
<tr>
<td>Polynomial interpolation, 219</td>
<td>Recursive filtering techniques, 20–21</td>
</tr>
<tr>
<td>Polyphase filters, 219</td>
<td>Reflection, 53</td>
</tr>
<tr>
<td>Poppet, Karl, 25, 302–3</td>
<td>Refraction, 53</td>
</tr>
<tr>
<td>Porat, Boaz, 231</td>
<td>Regions</td>
</tr>
<tr>
<td>Portable Network Graphics (PNG)</td>
<td>similarity and, 225</td>
</tr>
<tr>
<td>DEFLATE method and, 133</td>
<td>spatial proximity and, 225</td>
</tr>
<tr>
<td>differential coding, 150–51, 151t.12.1</td>
<td>Release, 47</td>
</tr>
<tr>
<td>Postscript, 118-19</td>
<td>Relevance feedback, 194, 196–98</td>
</tr>
<tr>
<td>Precision in classic information retrieval, 194</td>
<td>Remote controls, 110–11</td>
</tr>
<tr>
<td>Proncen, J.P., 172</td>
<td>Resolution</td>
</tr>
<tr>
<td>Printing press, 7, 13</td>
<td>cameras, 56</td>
</tr>
<tr>
<td>Privacy</td>
<td>of sensors, 32</td>
</tr>
<tr>
<td>future trends, 314–15</td>
<td>visual, 29</td>
</tr>
<tr>
<td>overview, 116–18</td>
<td>Responsiveness, 113</td>
</tr>
<tr>
<td>Processing unit, 96</td>
<td>Reverberation, 38</td>
</tr>
<tr>
<td>Proprioception, 15n.1</td>
<td>RGB color space, 61</td>
</tr>
<tr>
<td>Psychoacoustics</td>
<td>Riul, A., Jr., 34</td>
</tr>
<tr>
<td>absolute threshold of hearing (ATH), 165–66</td>
<td>RLE (Run-length encoding), 124–25</td>
</tr>
<tr>
<td>advanced perceptual compression</td>
<td>Robotics, 305n.1</td>
</tr>
<tr>
<td>and, 164–68</td>
<td>ROC/DET curve, 264</td>
</tr>
<tr>
<td>Bark scale, 165</td>
<td>Rock, Irvin, 25, 302</td>
</tr>
<tr>
<td>frequency/loudness resolution, 165–66</td>
<td>Rojas, Raul, 269</td>
</tr>
<tr>
<td>masking, 166–68</td>
<td>Roof edges, 226</td>
</tr>
<tr>
<td>Mel scale, 165</td>
<td>RTCP (Real Time Control Protocol), 102–3</td>
</tr>
<tr>
<td>Psychological effects, 119</td>
<td>RTP (Real Time Transport Protocol), 102–3, 171</td>
</tr>
<tr>
<td>Psychophysics, 236</td>
<td></td>
</tr>
<tr>
<td>Publishing formats for documents, 77</td>
<td></td>
</tr>
</tbody>
</table>
Index

RTSP (Real Time Streaming Protocol), 103
Run-length encoding (RLE), 124–25
Russell, Stuart J., 269
Rutherford, E., 121
Safety, 119–20
Salomon, David, 138
Sample level of computation, 210
Sampling, 209–10
Sax, John Godfrey, 4
Sayood, Khalid, 138
SB (Search buffer), 131
Scale-Invariant Feature Transformation (SIFT), 245–46, 282
Scherp, A., 25
Schuer recursion, 181
Schwartz Criterion, 279
Search buffer (SB), 131
Search engines, 297
Security, 118–19
Segmentation
split and merge, 228–30
video segmentation, 229–30
Selection of documents, 72
Semantic computing
future trends, 312–14
overview, 236
Semantic content-based retrieval, 202–3
Semantic gap, 22–23, 25, 292
Semantics, 15–16, 25
Semi-structured data, 191–92
Sensor integration, 86
Sensors
accelerometers, 30–31
audio sensors, 29
bias, 31
compasses, 30–31
defined, 28
digitization, 32–33
drift, 31–32
environmental sensors, 30–31
exotic sensors, 314
future trends, 308–9
human sensors, 28–29
hysteresis, 32
life sensors, 30
noise, 31–32
nonlinear behavior, 31–32
offset, 31
operational range, 31–32
overview, 28
properties of, 31–32
resolution, 32
saturation, 31
sensor, 29
smell sensors, 29
tactile sensors, 28–29
taste sensors, 29
temperature sensors, 30
time sensors, 31
types of, 28–31
weight sensors, 30
Sensory gap, 293
Sequential Minimal Optimization (SMO) algorithm, 257
Session Initiation Protocol (SIP), 103
Shannon, E., 127, 138
Shapiro, Linda, 231
Sharp, Helen, 121
Shirky, Clay, 79
SIFT (Scale-Invariant Feature Transformation), 245–46, 282
Sight sensors, 29
Signal processing
audio processing, 235
background and, 223–24
connected components and, 223
connectivity and, 223
downsampling, 217–18
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
dynamic range compression (DRC), 216–17
edges and, 226–28. (see also Edges)
figure-ground problem, 212
filter by example, 220–22
frames, 209
global level, 211
graphical filters, 222
holes and, 223–24
image processing, 235
interpolation filters, 218–19
levels of computation, 210–12
linear filters, 215–16
linear interpolation, 218
linear operators, 214–16
local level, 210–11
neighbors and, 222–23
non-linear filters, 216–19
object-background separation, 213
object level, 211
overview, 209, 230–31, 235
path and, 223
pixels, 209
polynomial interpolation, 219
polyphase filters, 219
quantization, 209–10
regions and, 225–26. (see also Regions)
sample level, 210
sampling, 209–10
spectral subtraction, 220
spline interpolation, 219
split and merge, 228–30
subsampling, 217–18
supersampling, 218
thresholding, 212–14
upsampling, 218
Index

Signal processing (cont.)
  variable gain amplifiers, 217
  video processing, 235
  video segmentation, 229–30
  VU meters, 222
  Wiener filter, 220–21
  SILK, 186
  Simon, J.D., 287
  Singhavi, L.M., 25
  Single Instruction, Multiple Data (SIMD), 211
  Sinha, P., 302
  SIP (Session Initiation Protocol), 103
  Skype SVOPC, 186
  Sliding windows, 131
  Smart phones
    event name tagging, 302
    identifying objects, 302
    identifying people, 300–2
    photo management, 300–2
  Smell sensors, 29
  Smee, J., 25
  SMIL (Synchronized Multimedia Integration Language), 79
  SMO (Sequential Minimal Optimization) algorithm, 257
  So, R.H.Y., 121
  Social media, 116
  Soft margin classification, 257
  Software evaluation through human subjects, 114
  Sommer, R., 121, 303
  Sony ATRAC, 165
  Sound
    adaptive concatenative synthesis, 48
    A-weighting scheme, 37
    concatenative synthesis, 48
    constructive interference, 38
    damage caused by, 120
    decibel scale, 37
    defined, 36
    destructive interference, 38–39
    echo, 38
    human-created sound, 45–46
    interference, 38
    microphones, 39–41. (see also Microphones)
    music, 46–48. (see also Music)
    overdriven signals, 41
    overview, 36
    physics of, 36–37
    production of, 43
    properties of, 38–39
    psychoacoustics. (see Psychoacoustics)
    recording of, 39
    reproduction of, 39, 41–43
    reverberation, 38
    speakers, 41–43
    speech, 43–45, 49. (see also Speech)
    synthesis of, 48
  wavetables, 48
  Sound amplitude quantization, 145–46
  Sound pressure, 37
  Source coding theorem, 127
  Spatial layout of documents, 75–76
  Spatial synchronization, 91–93
  Speaker diarization, 278–80
  cluster merging, 279
  re-segmentation, 279
  re-training, 279
  Speaker identification, 117
  Speaker recognition, 276–78
  evaluation metrics, 278
  general architecture, 277
  model generation, 277
  supervector generation, 277
  UBM index generation, 277
  Universal Background Model (UBM), 276–77
  Speakers, 41–43
  Specification of synchronization, 90–91
  Spectral subtraction, 220
  Speech, 43–45
    generally, 49
    approximants, 44
    compression. (see Speech compression)
    diphthongs, 44
    formants, 45
    fricative consonants, 44
    limited bandwidth, 45–46
    limited variance in harmonicity, 46
    limited volume, 46
    Lombard effect, 44
    monophthongs, 44
    phonemes, 44
    spectrogram, 45
    stop consonants, 43
    vowels, 44
  Speech activity detection, 272–73
  Speech analysis, 235
  Speech compression
    overview, 174
    CELP. (see Code Excited Linear Prediction (CELP))
    Global System for Mobile (GSM), 185–86
    Linear Protective Coding (LPC), 175–82. (see also Linear Protective Coding (LPC))
    properties of speech coders, 174–75
  Speech synthesis, 235
  Spline interpolation, 219
  Split and merge, 228–30
  Split attention, 84–86
  SQL (Structured Query Language), 191
  Static Huffman compressors, 131
  Steinmetz, Ralf, 93, 106
  Step discontinuities, 226
  Stiles, W.S., 64
  Stockman, George, 231
  Stoll, G., 172
Stop consonants, 43
Storage, 97
Storage and recording technology, 8
Stored video, 100
Streaming multimedia
application layer QoS control, 100
content replication, 101
continuous media distribution services, 100–1
IP Multicast, 101
media synchronization, 102
network filtering, 101
overview, 100, 106
protocols, 102–3
streaming servers, 101–2
Streaming servers, 101–2
String instruments, 47
Structure-based representation of documents, 77
Structured data, 191–92
Structured Query Language (SQL), 191
Subsampling, 217–18
Supersampling, 218
Support Vector Machines (SVMs), 255–57, 277, 281–82
Sustain, 47
Synchronization
content synchronization, 87–88
deadlines, 91
of documents, 76–77
explicit specification, 90–91
implicit specification, 90
intra-object specification, 91
layout frames, 92
levels of, 90
media synchronization, 102
multimedia streams, run-time support for, 90
overview, 82, 86–87, 93
spatial synchronization, 91–93
specification, 90–91
temporal synchronization, 88–90
time-dependent and time-independent media,
run-time support for, 90
Synchronized Multimedia Integration Language
(SMIL), 79
Systems
multimedia content analysis systems.
(see Multimedia content analysis systems
multimedia systems. (see Multimedia systems)
Tactile sensors, 28–29
Tag-based retrieval, 203–4
Tags, 24
Tamura image, 243
Task-orientation, 109–11
Taste sensors, 29
TCP (Transmission Control Protocol), 102
Telegraphy, 7
Telephony, 10
Television
cameras and, 56–57
communication by, 7–8
CRT technology, 58
differential coding and, 151
HDTV, 59, 124
lossy compression and, 141
speech activity detection, 273
visual quantization and, 147
Temperature sensors, 30
Temporal layout of documents, 76
Temporal modeling, 257–59
Temporal synchronization, 88–90
Tennebaum, J.M., 302
Terhardt, E., 172
Term frequency/inverse document frequency (TF/IDF), 195–96
Text editing, 72–73
“thinking outside in,” 109
3D video
cameras, 57, 59
damage caused by, 120
future trends, 310
privacy and, 314
Thresholding, 212–14
Tiling, 162
Timbre, 47
Time-based representation of documents, 77
Time sensors, 31
Tokens, 131
Torralba, Antonio, 246
Transistors, future trends, 307–8
Transmission Control Protocol (TCP), 102
Transparency, 112–13
Trash bins, 112
Turing completeness, 118–19
UBM (Universal Background Model), 281–82
UDP (User Datagram Protocol), 102
UmaSwami, 20
Uncertainty reduction, 84
Universal Background Model (UBM), 276–77, 281–82
Unstructured data, 191–92
Upsampling, 218
User Datagram Protocol (UDP), 102
User interface design principles
Desktop metaphor, 111, 112
dialog boxes, 110
false creativity, 111–12
fonts, 110
fostering learning, 112–14
graphical user interface (GUI), 108, 111–12
human-computer interaction (HCI), 108–9
menu bars, 108
questionnaires, 115
remote controls, 110–11
responsiveness, 113
### Index

User interface design principles (cont.)
- significance of testing, 116
- software evaluation through human subjects, 114
- task-orientation, 109–11
- “thinking outside in,” 109
- time of operations, 113
- transparency, 112–13
- trash bins, 112
- video surveillance tests, 114–15

Variable gain amplifiers, 217
- Vector quantization, 141–44
- k-means algorithm, 142–43
- linear quantization, 142
- x-means algorithm, 143–44
- Vectors, 227
- Ventriloquism, 83–84
- Verification vision, 296–97
- Video cameras, 56–57
- Video editing, 74
- Video processing, 235
- Video retrieval, 204–5
- Video segmentation, 229–30
- Video surveillance tests, 114–15
- Viola, P.A., 281
- Viruses, 119
- Visual dictionaries, 16–17
- Visual object recognition, 282–86
  - change detection, 283–84
  - difference pictures, 283–84
  - object tracking, 284–86
  - path coherence, 286
- Visual quantization, 146–47
- Visual texture, 242–43
  - coarseness, 242
  - contrast, 242
  - directionality, 243
  - Tamura image, 243
- Viterbi, Andrew, 259
- Viterbi algorithm, 259, 272–73

Vocal Tract Length Normalization (VTLN), 273–74
- Voiced/unvoiced detection, 176–77

Vectors, 227
- Weber-Fechner Law, 60, 145
- Web Image Search, 198
- Weight sensors, 30
- Welsh, T.A., 138
- WER (Word Error Rate), 264–66
- Westermann, G. Utz, 25–26
- What-You-See-Is-What-You-Get (WYSIWYG), 12
- Wiener, Norbert, 220
- Wiener filter, 220–21, 231
- Williams, Arthur B., 231
- Wilson, Jon S., 34
- Wind instruments, 47
- Winston, Patrick, 302
- Witten, I.H., 138, 269
- Word Error Rate (WER), 264–66
- Wu, Dapeng, 106

XML (Extensible Markup Language), 24, 191–92
- XMT (Extensible MPEG-4 Textual), 171

Yahoo, 297
- YouTube.com, 188, 234, 314
- YUV color space, 62

x-means algorithm, 143–44

XML (Extensible Markup Language), 24, 191–92
- XMT (Extensible MPEG-4 Textual), 171

Yahoo, 297
- YouTube.com, 188, 234, 314
- YUV color space, 62

Zero-Crossing Rate, 177
- Ziv, Jacob, 131, 138
- Zwicker, Eberhard, 165