

# Issues in Voice and Video Coding

Presented at ICME Panel I: Advances in Audio/Video Coding  
Technologies, July 12, 2011

**Jerry D. Gibson**

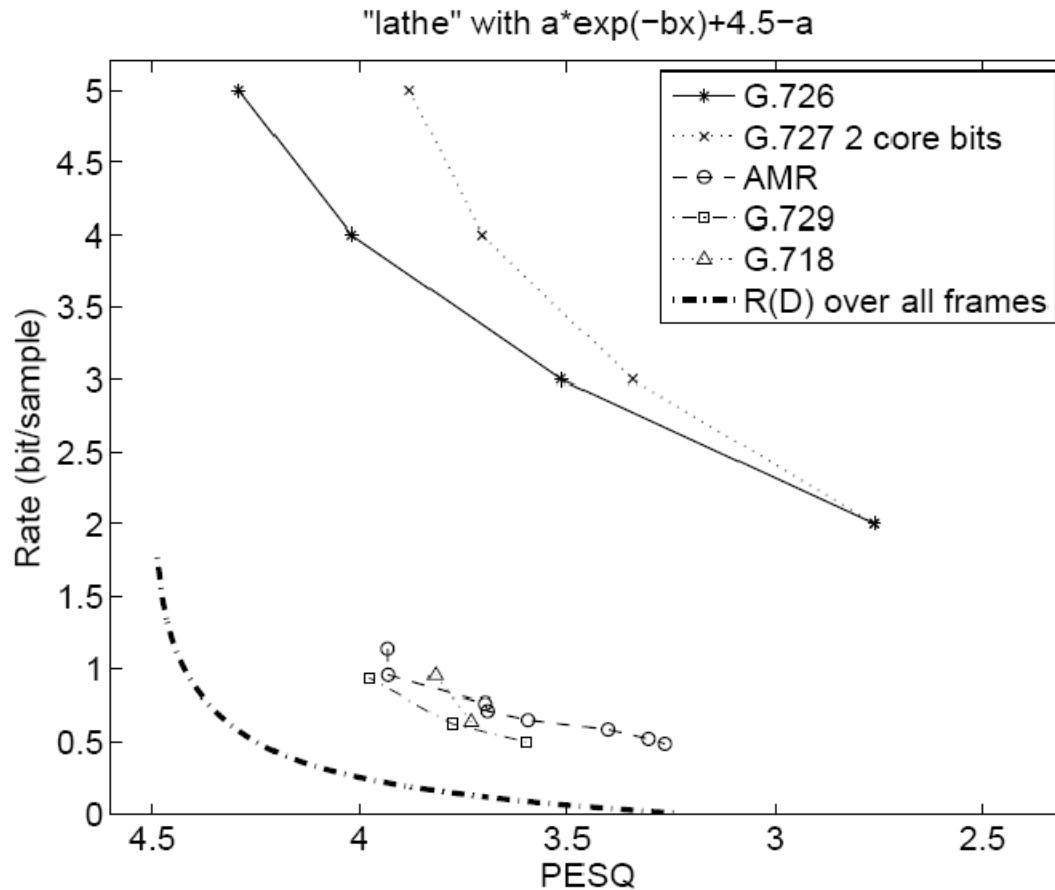
Department of Electrical and Computer Engineering  
University of California, Santa Barbara  
gibson@ece.ucsb.edu

This research has been supported by NSF Grant Nos. CCF-0728646  
and CCF-0917230

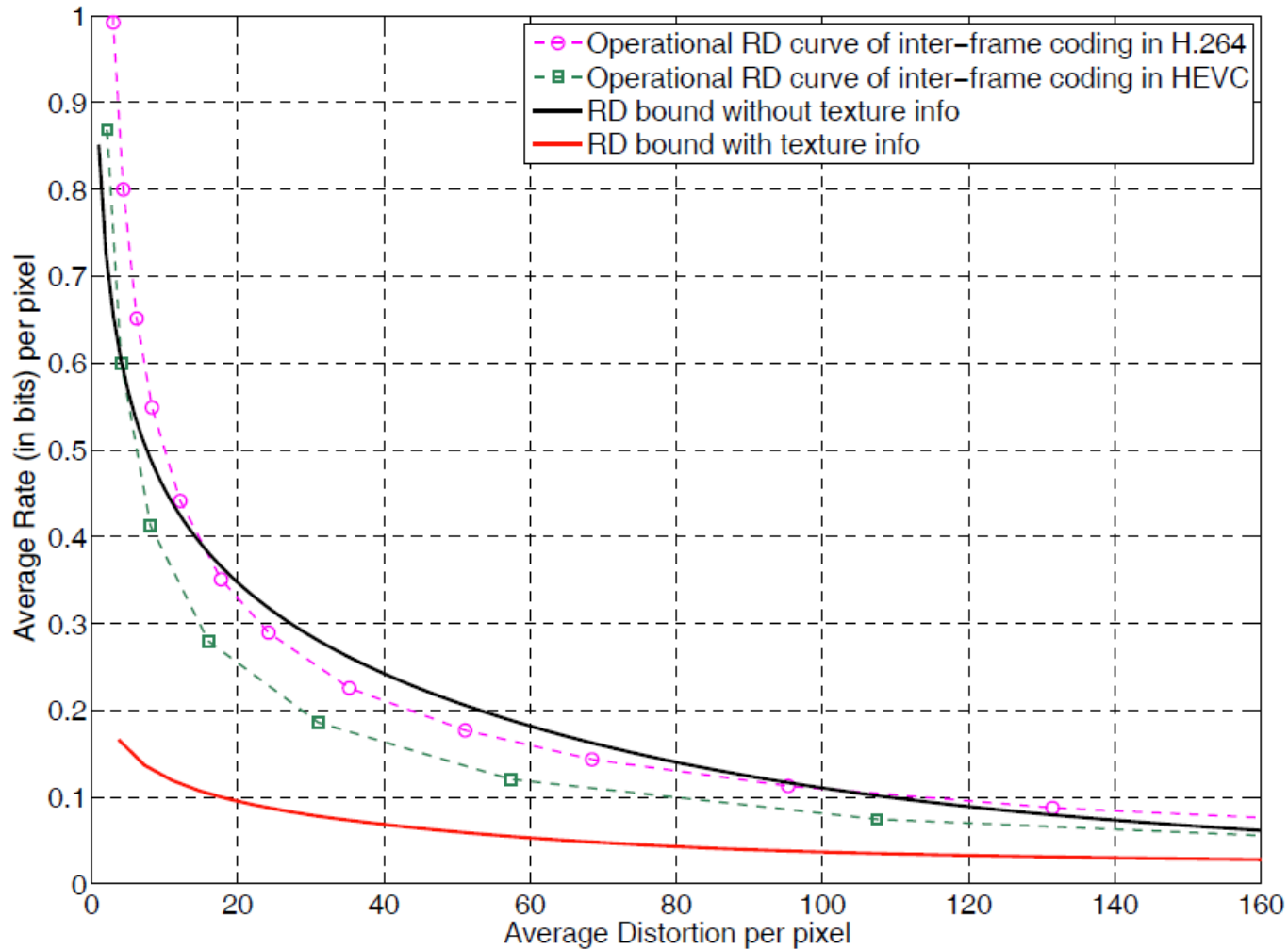
# Primary Considerations in Voice and Video Coding

- Rate
- Distortion
- Complexity
- Latency

# Bounding the Performance of Best Known Voice Codecs [1]



# Bounding the Performance of Best Known Video Codecs—Intra/Inter Mode [2-4]



# What About Latency and Complexity?

- Does Algorithmic Delay Deserve More Attention?
- Is Complexity Becoming Too Much?

# Narrowband (300-3400 Hz, 8 kHz sampling rate) speech codecs [5]

|   |             |                |              |  |
|---|-------------|----------------|--------------|--|
| Formal Name                                 | ITU-T G.711 | ITU-T G.726    | ITU-T G.729  | 3GPP AMR   |
| Technology                                  | Log PCM     | ADPCM          | CS-ACELP     | ACELP  |
| Bitrate(s)<br>(kbits/sec)                   | 48, 56, 64  | 16, 24, 32, 40 | 6.4, 8, 11.8 | 4.75, 5.15,<br>5.9, 6.7,<br>7.4, 7.95,<br>10.2, 12.2 |
| <b>Algorithmic<br/>Delay</b><br>(msec)      | 0.125       | 0.125          | 15           | 25   |
| <b>Comp.<br/>Complexity</b><br>(give units) | 0.01 MIPS   | 1.25 MIPS      | 18 MIPS      | 11.9-16.7<br>WMOPS                                   |

# Wideband Speech Codecs [5]

|                                     |                   |               |   |  |  |
|-------------------------------------|-------------------|---------------|---|--|--|
| Formal Name                         | ITU-T G.722       | ITU-T G.722.1 | ITU-T G.722.2<br>3GPP AMR-WB  | ITU-T G.718  | ITU-T G.719  |
| Technology                          | Sub-band<br>ADPCM | MLT           | ACELP   | ACELP, MDCT  | Adaptive<br>resolution<br>MDCT, FLVQ   |
| Audio<br>Bandwidth(Hz)              | 50-7000           | 50-7000       | 50-7000   | 50-7000  | 20-20000   |
| Bitrate(s)<br>(kbits/sec)           | 48, 56, 64        | 24, 32        | 6.6, 8.85, 12.65,<br>14.25, 15.85,<br>18.25, 19.85,<br>23.05, 23.85 | 8,12,16,24,32 &<br>12.65 (G.722.2,<br>AMR-WB,<br>VMR-WB<br>Interop Mode) | 32...128 steps<br>of 4 kbps up to<br>96 kbps, steps<br>of 8 kbps up to<br>128 kbps |
| <b>Algorithmic<br/>Delay (msec)</b> | 1.625             | 40            | 25  | 32.875 to<br>43.875  | 40   |
| <b>Comp.<br/>Complexity</b>         | 10 MIPS           | < 5.5 WMOPS   | 27.2-39.0<br>WMOPS  | 57 WMOPS   | 15.39 - 21<br>WMOPS  |

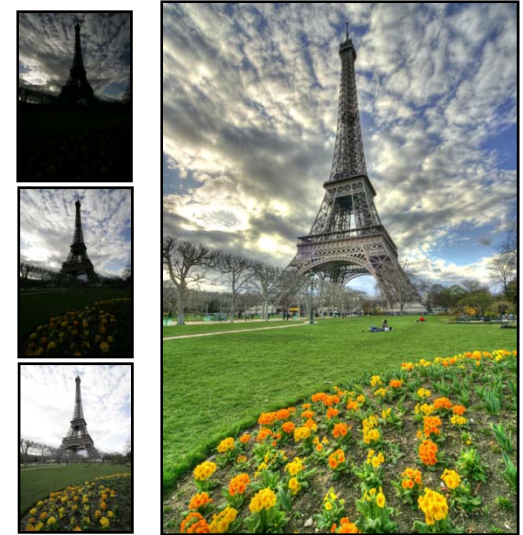
# Current Efforts for Voice/Audio Coding: Quality of User Experience

- Wider Bandwidths
  - Narrowband (200 to 3400 Hz)
  - Wideband (50 Hz to 7 kHz)
  - Superwideband (50 Hz to 14 kHz)
  - Fullband (20 Hz to 20 kHz)
- Stereo
- Spatial Localization
- Multiparty Calls
- Acoustic and Background Noise

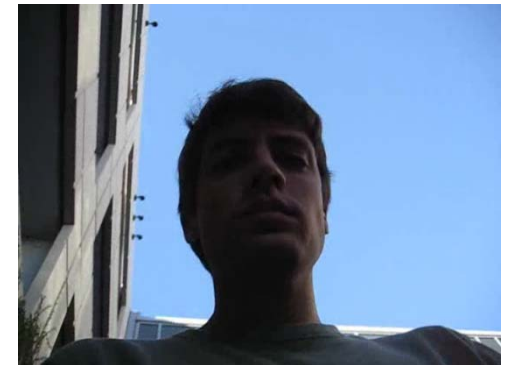
# High Dynamic Range Video for Handhelds

- Inexpensive video cameras have limited dynamic range – saturated pixels [6]
- HDR photography combines multiple exposures, yet we **need new methods for video** [7]
- Applications:
  - Videoconferencing
    - Saturated pixels on user's face hurt experience
    - Mobile/Handhelds: extreme outdoor lighting conditions
  - Security/Surveillance [8]
    - Dynamic range crucial to “see” environment
    - Temporal fidelity secondary
  - Need low-cost solution (<\$10)

© 2006 Jacques Joffre



**HDR Still Photography**



**Mobile Videoconferencing  
(poor lighting!)**

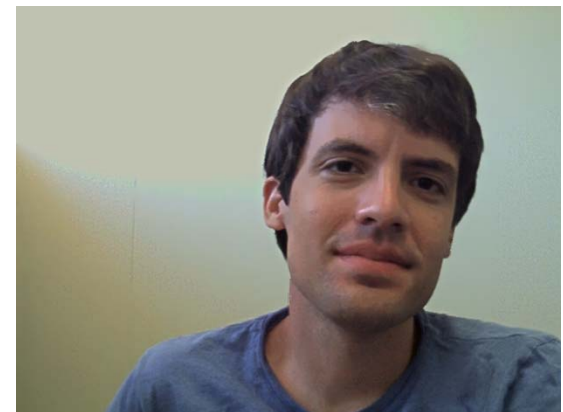
# Recent Results on High Dynamic Range Video for Handhelds

- Alternate between short/long exposures
- Combine adjacent frames to achieve HDR at the **same frame rate**
- Need to **remove ghosting** with motion compensation and filtering [9]

Low Dynamic Range Inputs



High Dynamic Range Outputs



# Viewing and Sensing 3D Video on Handhelds

- Glasses-free autostereoscopic displays now available on handheld gaming devices and phones
- Back-facing stereo cameras are standard
- Front-facing stereo cameras – **3D Videoconferencing**
- 3D can enhance experience if done correctly

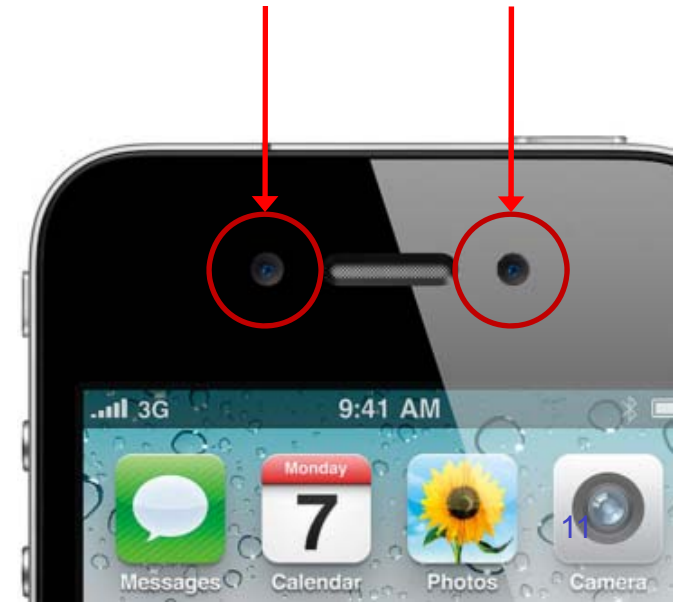


HTC EVO 3D



LG Thrill

Front-facing Stereo Camera?



# Issues for Handheld 3D Videoconferencing

- How to achieve effective and comfortable 3D for video communications on handhelds
- Close-up stereo photography is notoriously difficult! [10-11]
  - Optimal camera placement for display and analysis not the same
    - Need small stereo baseline (~9mm!) to reduce disparities
    - Need wider baseline for significant depth reconstruction
  - Need to **adjust disparities in real-time** according to scene depth [12]
- **Combine 3D and HDR**



Handheld 3D Videoconferencing



Nintendo 3DS "Depth Slider"

# References

- [1] Y.-Y. Li and J. D. Gibson, "Rate Distortion Bounds for Speech Coding Based on A Perceptual Distortion Measure (PESQ-MOS)," IEEE International Conference on Multimedia and Expo (ICME), Barcelona, July 2011.
- [2] Hu, J.; Gibson, J.D.; , "New rate distortion bounds for natural videos based on a texture dependent correlation model in the spatial-temporal domain," 46th Annual Allerton Conf. on Communication, Control, and Computing, pp.996-1003, 23-26 Sept. 2008
- [3] Hu, J.; Gibson, J.D.; , "Rate distortion bounds for blocking and intra-frame prediction in videos," 43rd Asilomar Conf. on Signals, Systems and Computers, pp.573-577, 1-4 Nov. 2009
- [4] Wiegand, T.; Han, W.-J.; Bross, B.; Ohm, J.-R.; Sullivan, G.J.; "WD2: Working Draft 2 of High-Efficiency Video Coding," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, doc. JCTVC-D503, Daegu, South Korea, 20–28 Jan. 2011.
- [5] ITU-T SG16 Q7/16, "Media Coding Summary Database," Nov. 2009.

# References

- [6] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005.
- [7] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, “High dynamic range video,” in ACM SIGGRAPH, New York, NY, USA, 2003, pp. 319–325.
- [8] S. Mangiat and J. Gibson, Inexpensive High Dynamic Range Video for Large Scale Security and Surveillance, MILCOM, Baltimore, MD, Nov 2011.
- [9] S. Mangiat and J. Gibson, “High dynamic range video with ghost removal,” in SPIE Optical Engineering & Applications, 2010.
- [10] L. Lipton, Foundations of the stereoscopic cinema: a study in depth. Van Nostrand Reinhold, 1982.
- [11] B. Mendiburu, 3D Movie Making: Stereoscopic Digital Cinema from Script to Screen. Focal Press, 2009.
- [12] Manuel Lang, Alexander Hornung, Oliver Wang, Steven Poulakos, Aljoscha Smolic, and Markus Gross, “Nonlinear disparity mapping for stereoscopic 3d,” ACM Trans. Graph., vol. 29, no. 3, pp. 10, 2010.