Squeeze Play:
The State of Ady0 Cmprshn

Scott Selfon
Senior Development Lead
Xbox Advanced Technology Group | Microsoft

GAME DEVELOPERS CONFERENCE 2012
SAN FRANCISCO, CA
MARCH 5-9, 2012
EXPO DATES: MARCH 7-9

Agenda

- Why compress?
- The tools at present
- Measuring success
- A glimpse of the future

The Philosophy of Compression

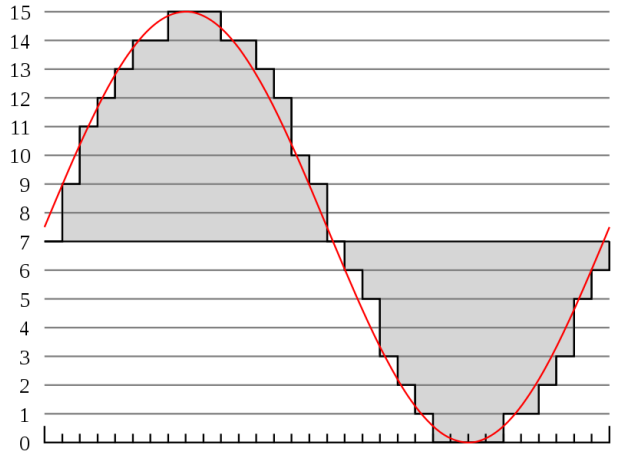
The tools of the present

- Black box codecs
  - Parameters that may or may not have well-understood meaning
  - Results that may or may not be appropriate
- Compression targets
  - Iteration slow enough to be discouraged
  - Bulk quality settings

Compression formats, ca. 2012

- Lossless codecs (<3:1): FLAC, Apple Lossless
- Lossy codecs
  - “Reductions” (up to ∞:1): sample rate, bit depth, channel count, noise floor, culling
  - Time domain: A-law/u-law, ADPCM (~4:1)
  - Perceptual (6-40+:1): MP3, Ogg Vorbis, XMA, etc.
- Hybrids (vary): AAL, WavPack, MP3 variants

PCM — Yes, still compression!



- Pulse Code Modulation
  - Analog signal regularly sampled and stored digitally
- Bit depth: Storage representation of a sample
  - Linear PCM = linear quantization
- Sampling rate: Frequency of analog signal capture or reproduction
  - Nyquist frequency (SR/2)

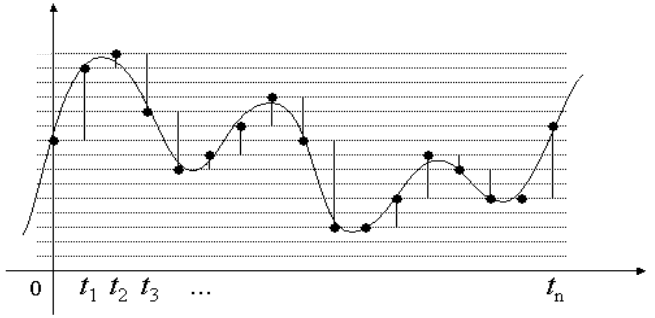
PCM and Quantization

- Frequency quantization
  - 44,100 Hz can represent sound frequencies up to 22,050 Hz
- Amplitude quantization
  - 16 bits: $20 \log \frac{2^{16}}{2} =$ ~90 dB range
  - 8 bits: $20 \log \frac{2^{8}}{2} =$ ~42 dB range

PCM A-Law/μ-Law (G.711)

- Pulse Code Modulation (1972, ITU 1988)
- Adds *companding* support
  - A-Law (13 bit signed→8 bit signed)
  - μ-Law (14 bit signed→8 bit signed)
  - Encodes location of most significant non-zero bit, drops one or more LSBs
  - Designed for telephony (8 kHz, 8 bit)

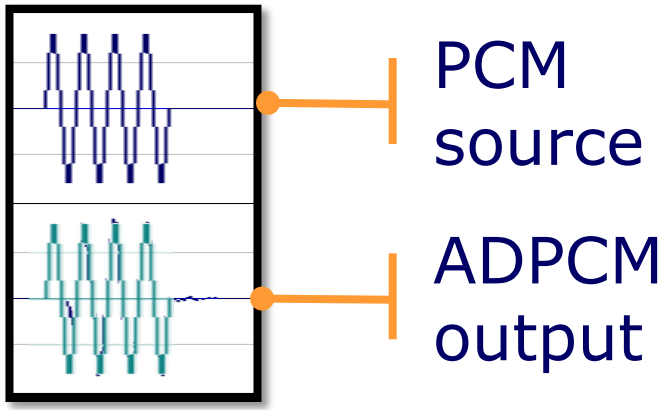
ADPCM (G.726)



- Adaptive Differential Pulse Code Modulation (ITU 1970s, IMA 1990s)
  - Stores difference between samples
  - Quantized to a step size lookup table
  - ~4:1 compression (16 bits→4 bits)
- Cheap to decode on CPU, straightforward to HW accelerate

ADPCM Artifacts

- Codec assumption: Signal slope doesn't change suddenly
  - Poor response to transients, quick attacks
  - Settling time before silence
  - Challenged particularly at lower sampling rates (<32 kHz)
- Step size quantization errors

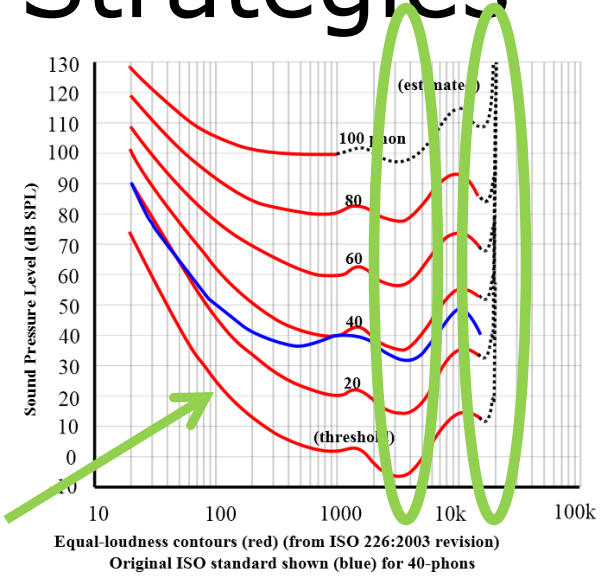


Perceptual Compression

- MP3, WMA, XMA, AAC, Ogg Vorbis, ATRAC, AC-3…
- Psychoacoustic: based on human frequency sensitivities
  - Frequency-domain compression
  - Take advantage of limits of auditory perception

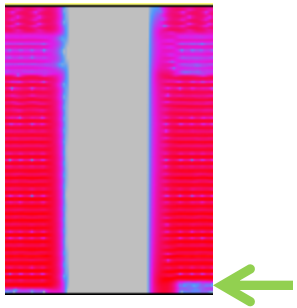
Perceptual Compression Strategies

- Frequency sensitivities
  - Nominally 20 kHz, often realistically 16 kHz
  - Most sensitive to speech range
- Absolute threshold of hearing
- Masking

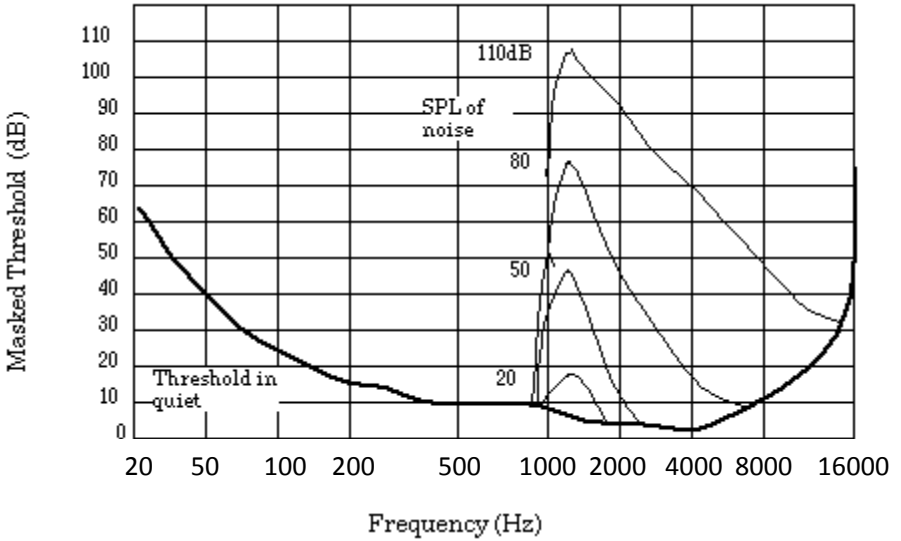


Acoustic Masking

- Frequency Masking



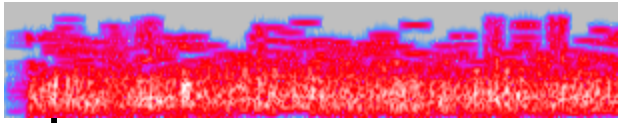
- Time Masking
  - Forward masking
  - Backward masking



A narrow 1200 Hz noise band masks sounds at higher frequencies (Scharf 1975)

Perceptual Codec Artifacts

- Time ⇆ frequency domain artifacts
  - Window size limits accuracy for transients: ringing or pre-echoes
  - Loss of phase information: warbles, ‘underwater’



- Channel collapse/recreation artifacts
  - Spatial loss and cross-talk

Game-Specific Perceptual Artifacts

(Or, Games are from Mars, Codecs are from Venus)

- Pitch shifting
- Mixing / Synchronization
- Repetition and Reuse
- Looping

New Dog, Old Tricks

- Sample rate reduction
- Bit depth reduction
- Channel reduction
- Normalization

…can all be less effective (or ineffective) with perceptual codecs

Choosing a Compression Format

- Support (device platform, middleware)
- Performance tradeoffs (CPU or hardware)
- Licensing (or lack thereof)

Evaluating Codec Capabilities

- Storage and bandwidth
- Decode latency
- Multichannel support (and leveraging)
- Looping accuracy
- Seamless seeking
- Perceptual quality

Measuring Success

- Critical listening and perceptual codecs

Squeeze Play: The Game Show

Which wave is more compressed?

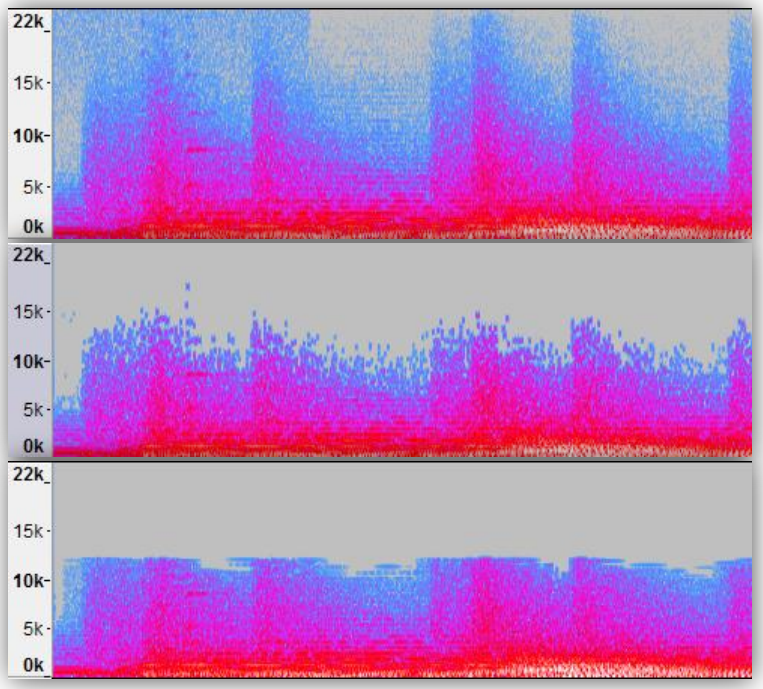
A	B	C
PCM (46 KB)	XMA q60 (8 KB, ~6:1)	ADPCM (12.5 KB, ~3.6:1)

Which wave is more compressed?

Input (44.1 kHz PCM)
1.85 MB

A Output (XMA, quality 1)
140 KB [13:1 compression]

B Output (xWMA, 48 kbps)
76 KB [24:1 compression]

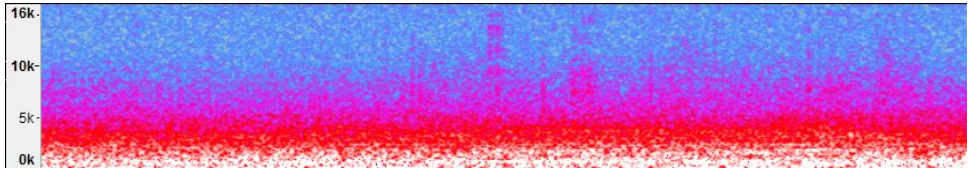


Measuring Success

- Critical listening and perceptual codecs
- Visual evaluations

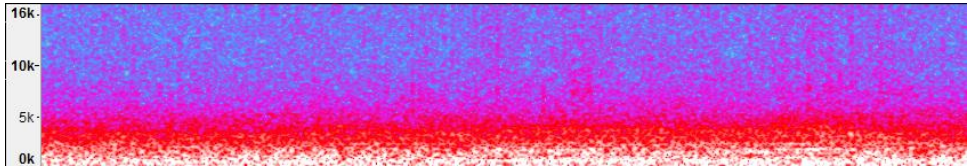
Which wave is more compressed?

Input (32 kHz PCM)
298 KB



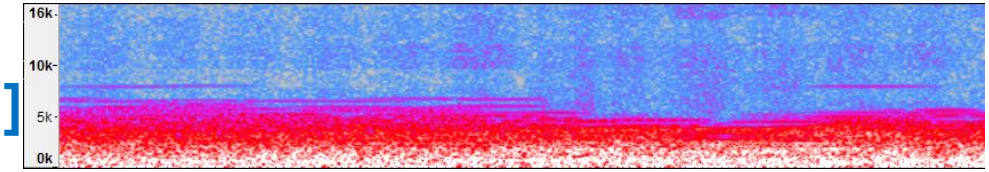
A

Output (ADPCM)
82 KB [3.6:1 compression]



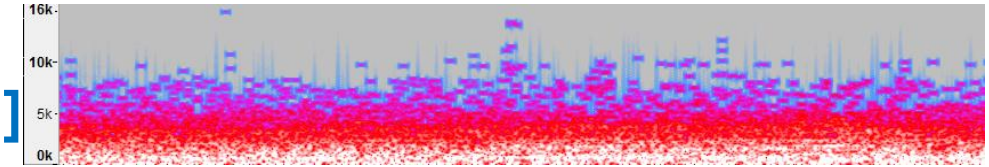
B

Output (xWMA, 20 kbps)
16 KB [18.6:1 compression]



C

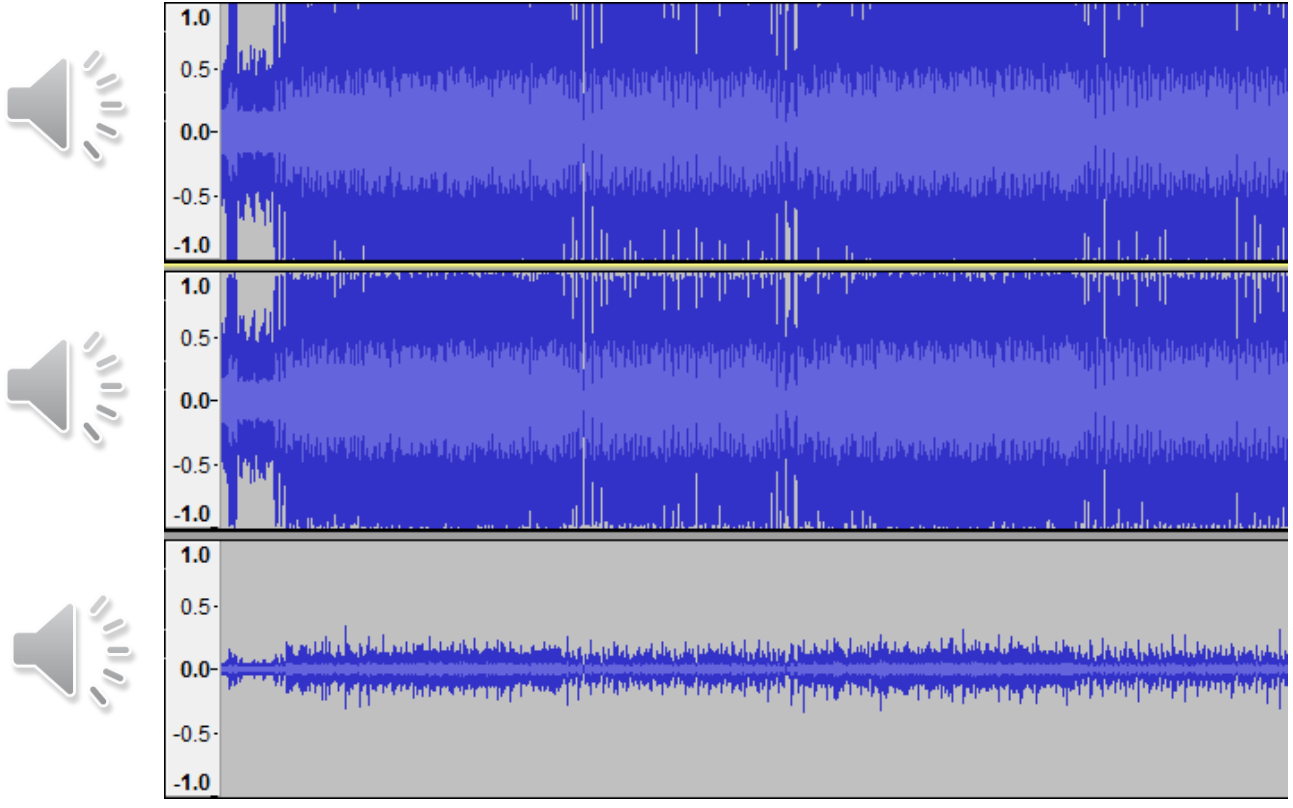
Output (XMA, quality 1)
28 KB [10.6:1 compression]



Measuring Success

- Critical listening and perceptual codecs
- Visual evaluations
- Delta evaluations (Taylor, 2011)

Delta Evaluations

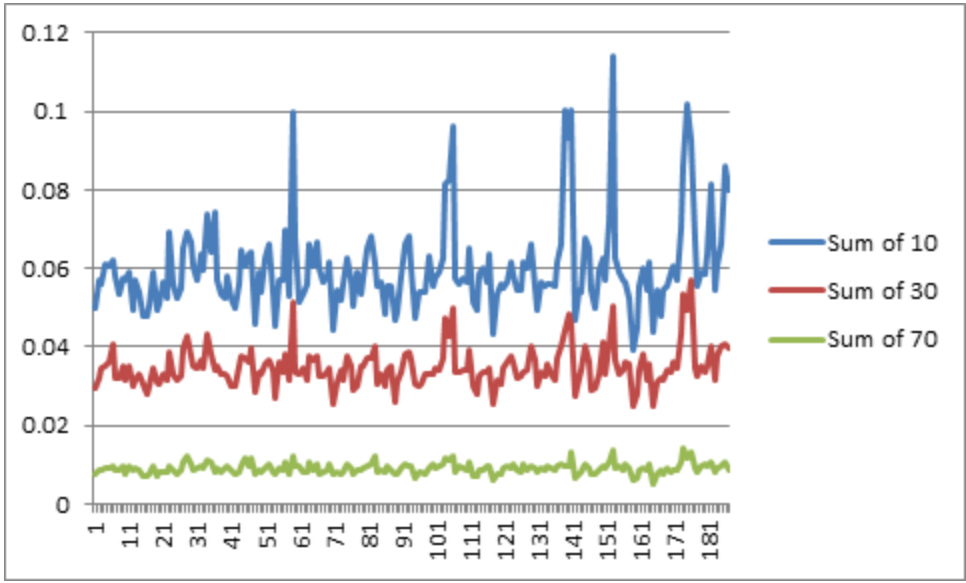


Measuring Success

- Critical listening and perceptual codecs
- Visual evaluations
- Delta evaluations (Taylor, 2011)
- Automated evaluation
  - PESQ/POLQA (ITU-T Rec. P.863)
  - PEAQ (ITU BS.1387-1)
  - Noise to Mask Ratio (NMR)

NMR Evaluation

- Noise to Mask Ratio
  - Windowed evaluation of Signal-to-Mask Ratio (SMR) minus Signal-to-Noise Ratio (SNR)



NMR at three XMA quality settings (Mathews 2012)

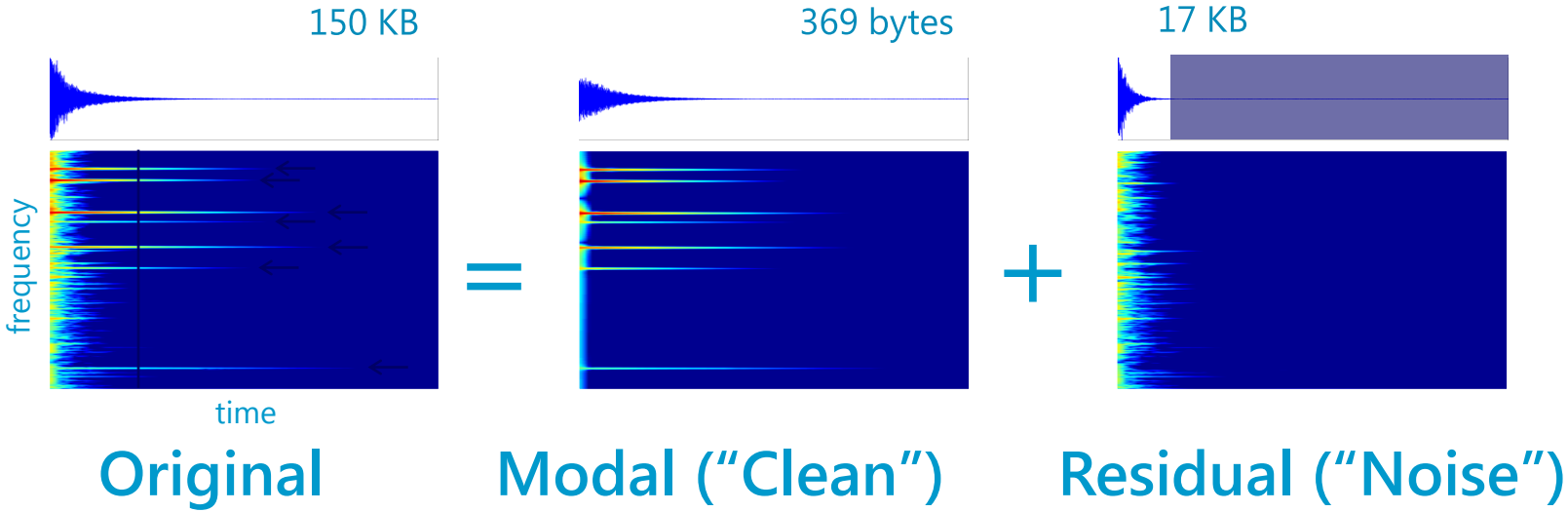
The Compression of the Future?

- Self-correcting/adjusting compression
- Communicating more with less
  - Linguistic sounds and speech synthesis
  - MIDI music: the revenge?
  - Parameterized procedural synthesis
- Case study: impacts

Impacts

- Resonant decay + transient
- Compress as modes + residual (>150:1)

Lloyd, Raghuvanshi, Govindaraju (ACM, 2011)



Conclusions

- Know thy artifacts
  - And use appropriate techniques to counter
- What's the playback context?
- More robust qualitative evaluation
  - Avoid the 'bulk' knob
  - Consider automating listening tests

Questions?

scottsel@microsoft.com

Xbox LIVE Gamertag: Timmmmmay