1980

clarity, durability

1995

Miniaturization,
data compression

2010

Psychoacoustics and *two kinds of* perceptual audio coding

Duality between data compression and Watermarking

Yi-Wen Liu, Ph.D.

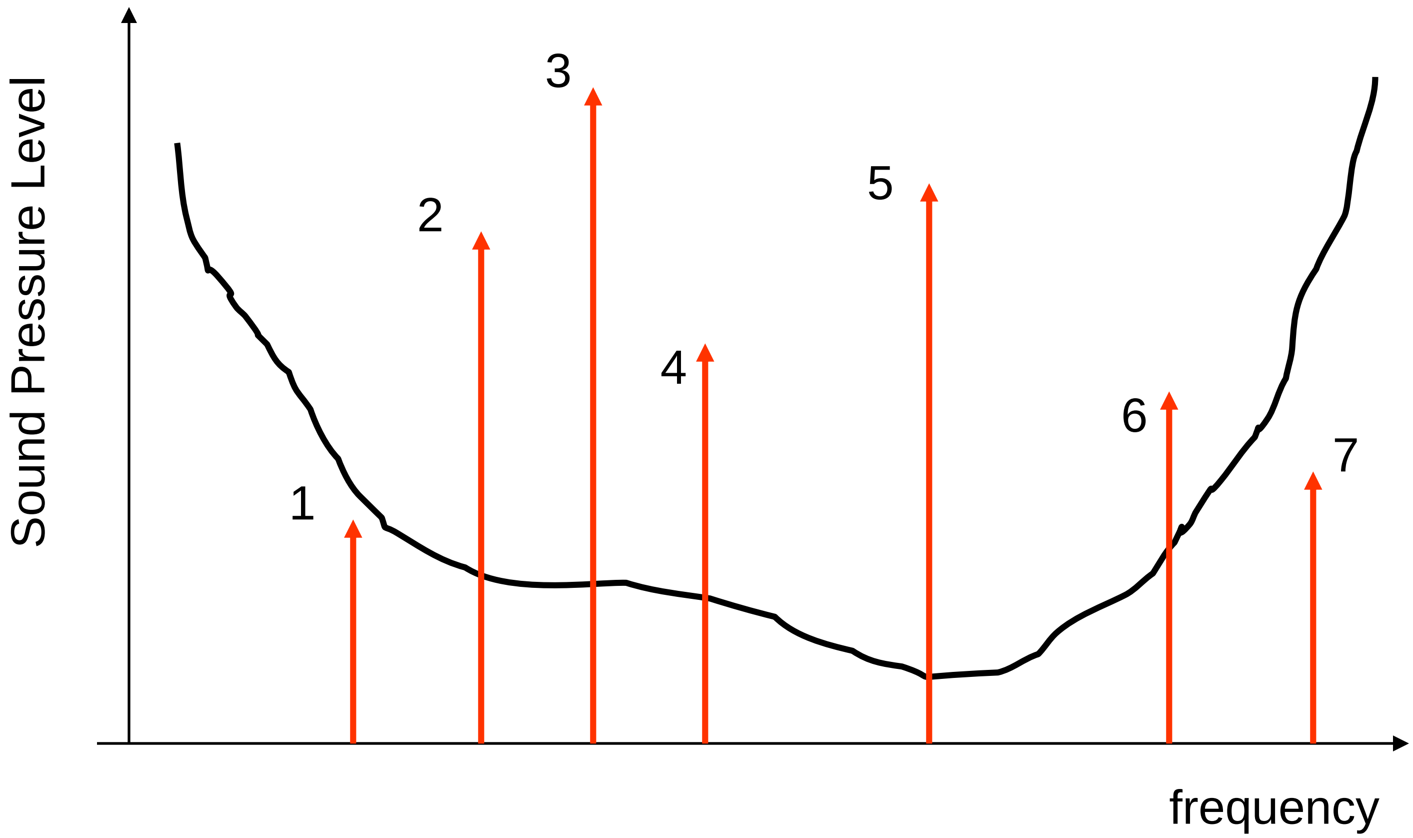
EE6641: Analysis and Synthesis of Audio Signals
Nov. 30, 2010
Last updated Dec. 21, 2015

Outline

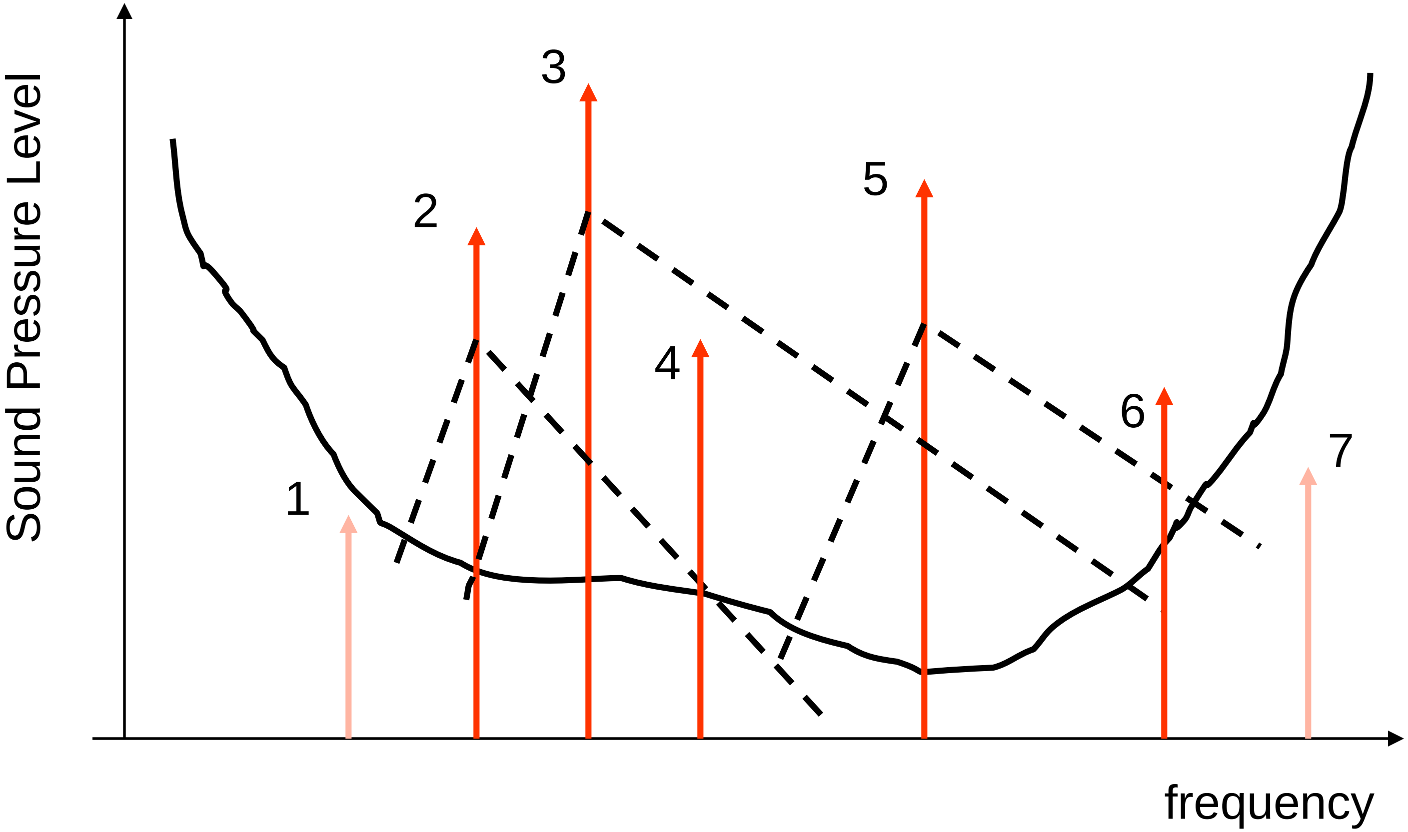
- Data Compression
  - Psychoacoustic modeling in mp3, iTune, etc.
  - Quantization and round-off error
  - Noise hiding principle

- Digital Watermarking
  - Applications
  - Dual to data compression
  - Games between embedder and attackers
  - Useful for *digital right management*?

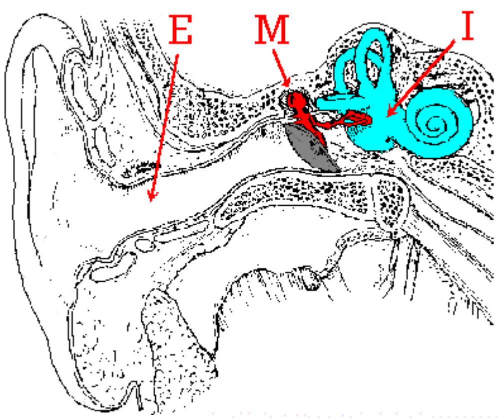
MP3 removes the in-audible,…

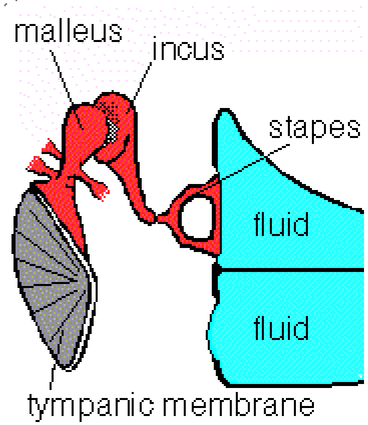


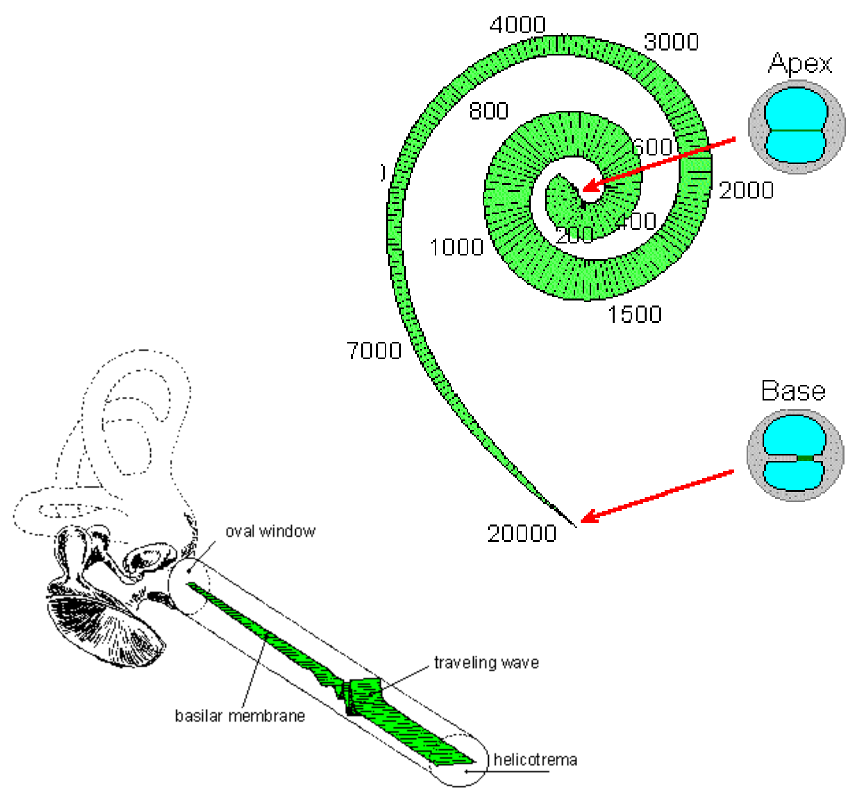
and also considers psychoacoustic *masking*



Basilar membrane as a spectrum analyzer



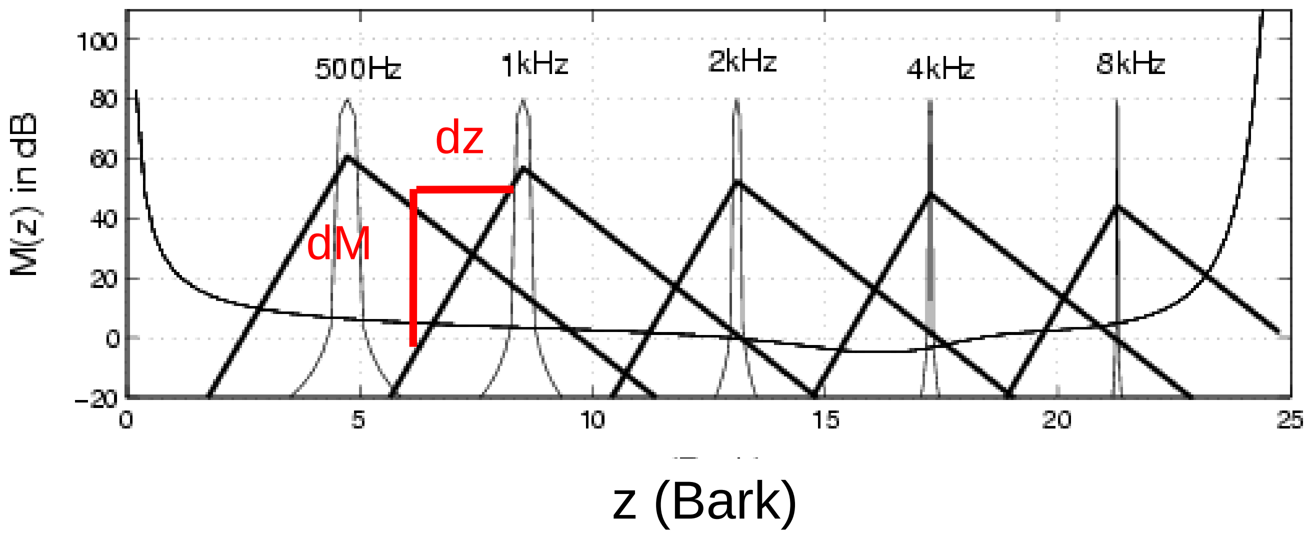




Images downloaded from:
http://www.vimm.it/cochlea/cochleapages/theory/

Spatial excitation of pure tones along basilar membrane is approximately a constant, triangular shape.

- *M(z)*, called *spreading function*,
  - *dM/dz* ≈ **27dB/Bark** (or 1dB / 50 µm) on the low-frequency side.
- Frequency z is measured in Barks.
  - <u>Δz = 1 (Bark)</u> corresponds to ~<u>1.3mm</u> on basilar membrane



More about masking

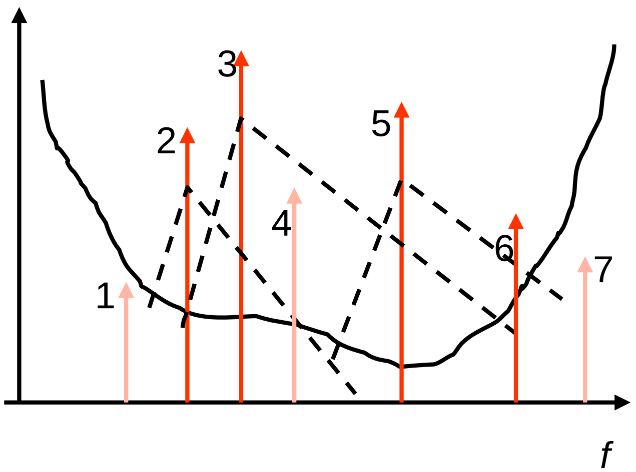
- Tonal vs. noise masker
- Forward masking
- Pre-masking

Is audio data compression all about removing the inaudible?

- In this example, 3 out of 7 tones are removed. So, compression ratio is approximately 42%.

- To compare,
  - MP3: >80% (5:1),
    *AAC**: > 90% (10:1),

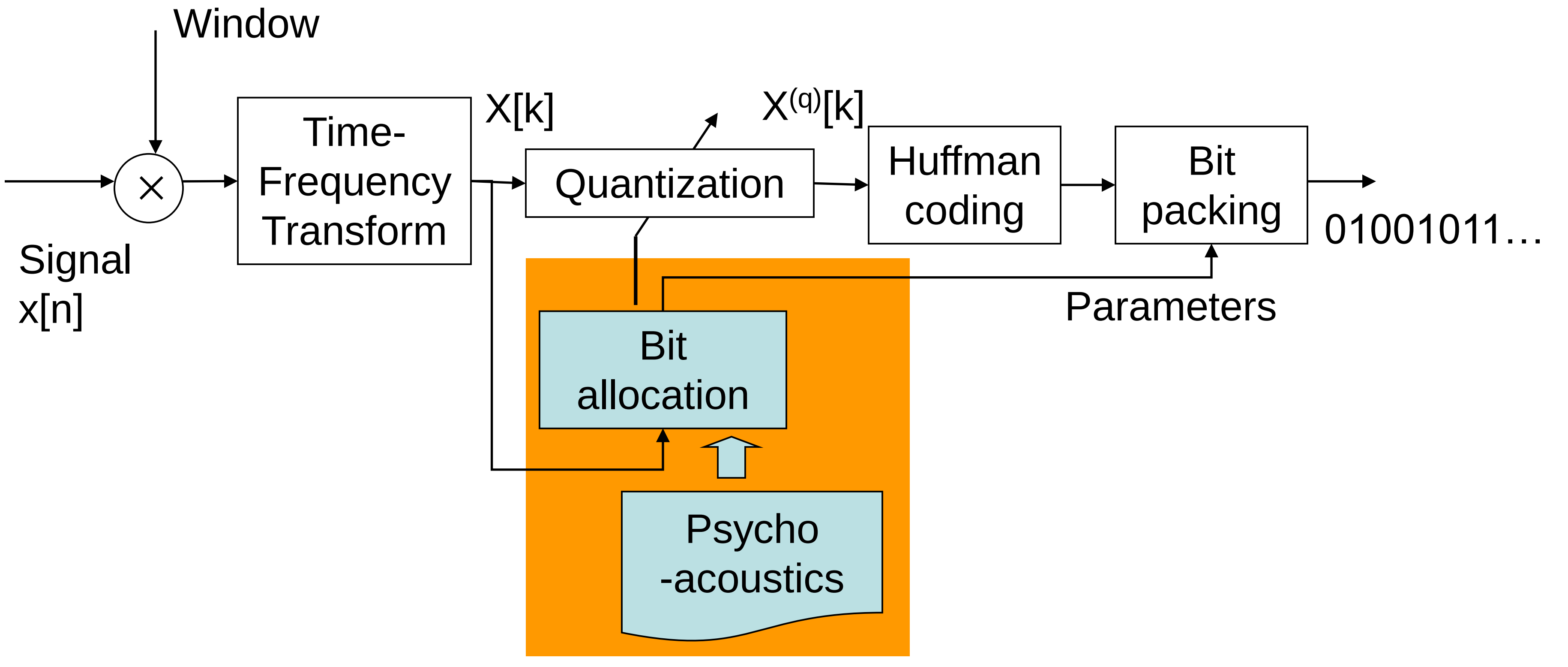
  with "good" sound quality.



*AAC is Apple's audio format.

Noise shaping in audio coding:

Quantization error $X^{(q)}[k] - X[k]$ to be masked



Binary Representation of Numbers

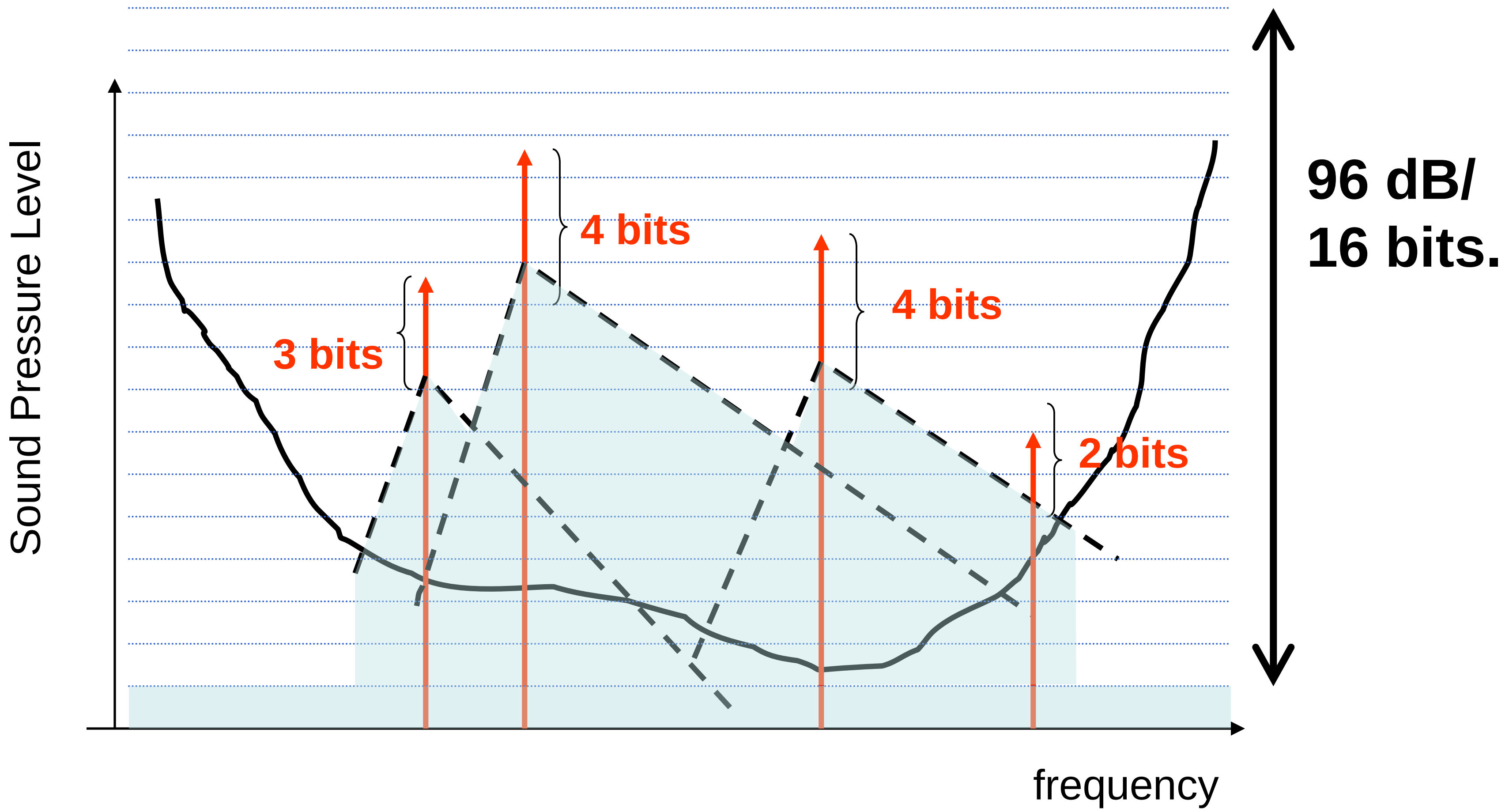
- Step 1: Determining number of bits (N) to use
- Step 2: Divide full signal range $\pm$R by 2^N equally spaced *quantization points* (QP).
  - Each QP uniquely represented by binary string of length N.
- Step 3: For any floating-point number, find nearest QP.
  - R = 32, N = 6, $2^N = 64$;
    so, $(20.79)_{10} \sim 21 = (010\ 101)_2$

More on quantization

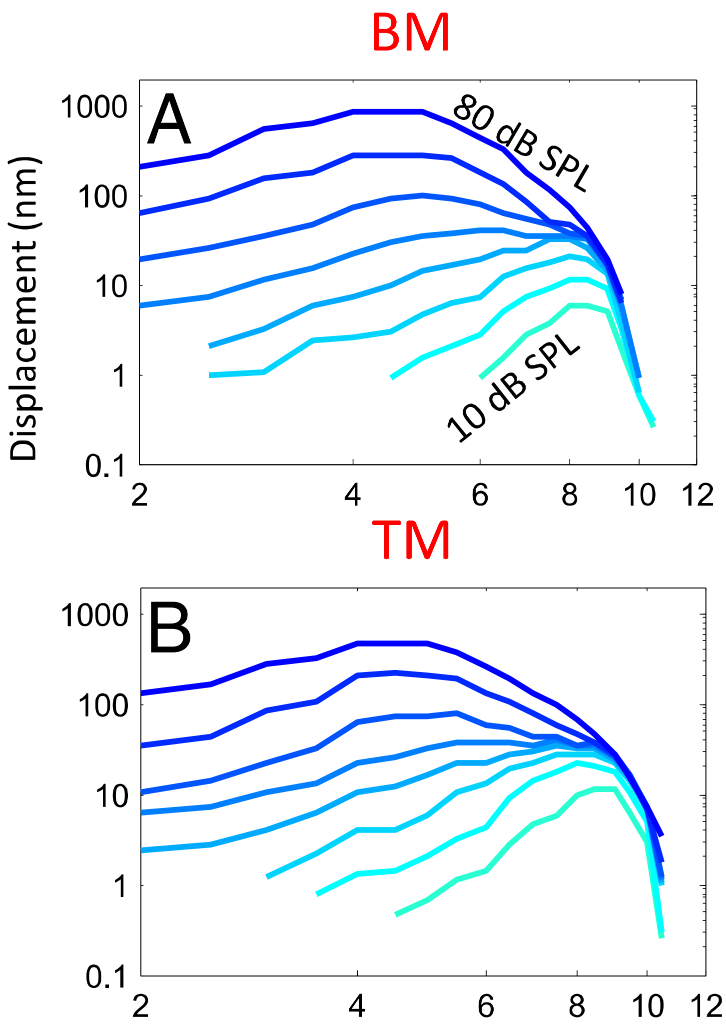
Range	N	Quantiz. Pts.	Step-size	Decimal	Approx.	Binary
±32	6	64	1	20.79	21	010 101
±32	5	32	2	20.79	20	010 10
±32	4	16	4	20.79	20	010 1
±32	3	8	8	20.79	24	011

Statistically, each additional bit improves signal-to-noise ratio (SNR) by 2-fold.

SNR = 6 dB/bit

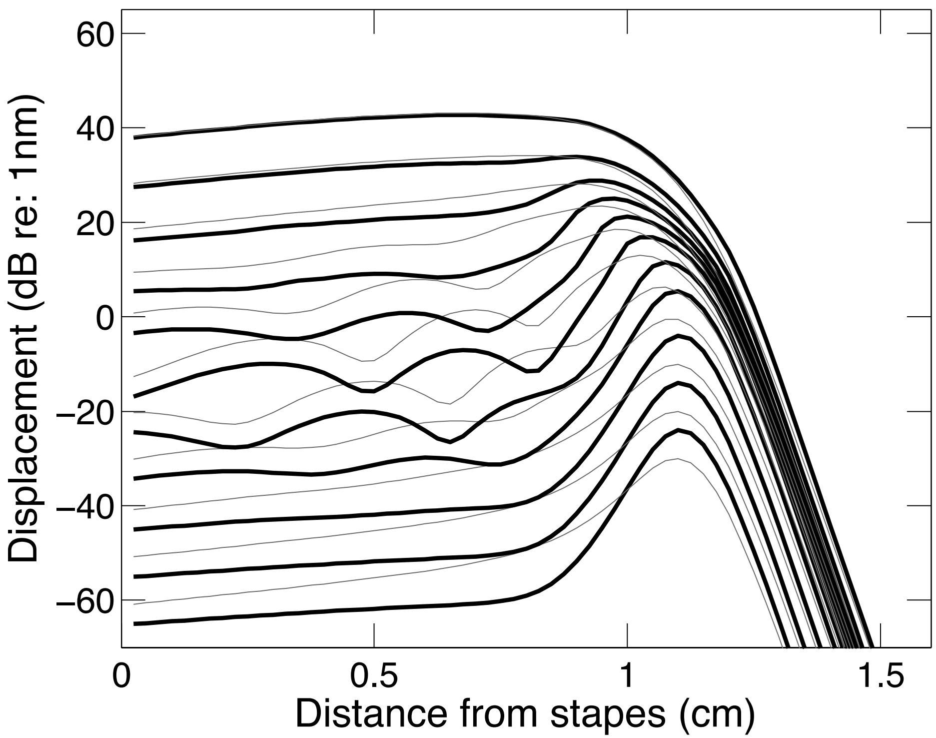

Sound Pressure Level
3 bits
4 bits
4 bits
2 bits
96 dB/
16 bits.
frequency

Remarks on the excitation pattern

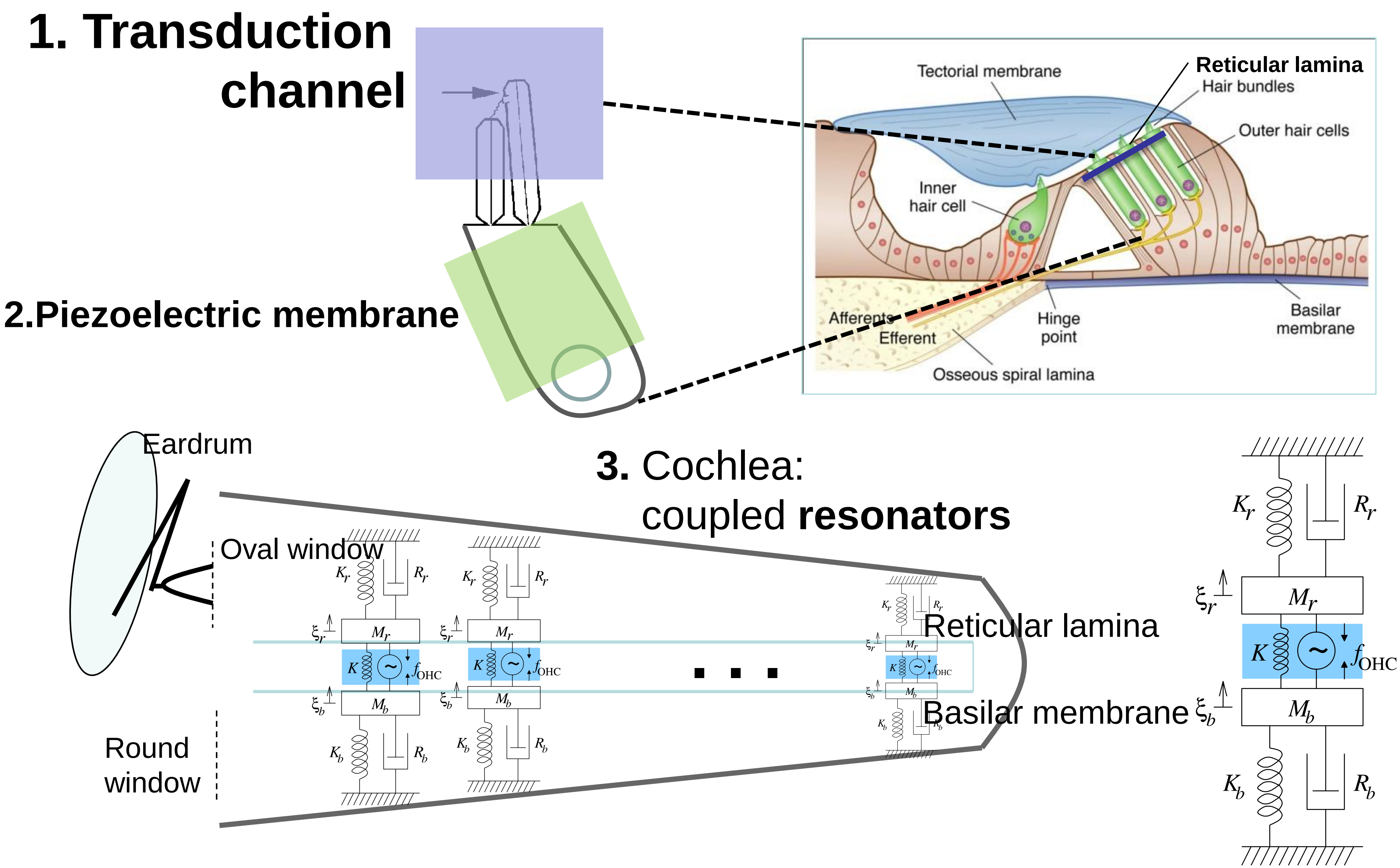


Lee et al. (2015)

Liu and Neely (2010)



電生理、細胞組織結構、與流體方程之整合模型 (Liu and Neely, 2010)



Work done at Boys Town, and continuing at Tsinghua (Hsinchu, Taiwan)

- Liu, Y.-W., Yu, L.-M., Wu, P.-J. (2014). “Close-loop simulation of the medial olivocochlear (MOC) anti-masking effects,” to appear in *Proc. the 12th Int. Mechanics of Hearing Workshop*, Attica, Greece.
- Liu, Y.-W. (2014). “Stationary noise responses in a nonlinear model of cochlear mechanics: Iterative solutions in the frequency domain,” submitted to *J. Acoust. Soc. Amer.*
- Liu, Y.-W., and Neely, S.T. (2013). “Suppression tuning of distortion-product otoacoustic emissions: Results from cochlear mechanics simulation,” *J. Acoust. Soc. Am*. **133**, 951-961.
- Liu, Y.-W. and Neely, S.T. (2012a). “Quasilinear cochlear responses to noise can result from instantaneous nonlinearities,” in *What Fire is in Mine Ears: Progress in Auditory Biomechanics*, C. A. Shera and E. Olson (Eds.). New York: American Institutes of Physics, pp. 218-223.
- Liu, Y.-W., and Neely, S. T. (2010). “Distortion product emissions from a cochlear model with nonlinear mechanoelectrical transduction in outer hair cells,” *J. Acoust. Soc. Am*. **127**, 2420-2432.
- Keefe D. H., Fitzpatrick D., Liu, Y.-W., Sanford C. A., and Gorga, M. P. (2010). “Wideband acoustic reflex test in a test battery to predict middle-ear dysfunction,” *Hear. Res.* **263**, 52-65.
- Liu, Y.-W., and Neely, S.T. (2009). “Outer hair cell electromechanical properties in a nonlinear piezoelectric model,” *J. Acoust. Soc. Am*. **126**, 751-761.
- Sanford, C.A., Keefe, D.H., Liu, Y.-W., Fitzpatrick, D.F., McCreery, R.W., Lewis, D.E., and Gorga, M.P. (2009) “Sound-conductance effects on DPOAE screening outcomes in newborn infants: Test performance of wideband acoustic transfer functions and 1-kHz tympanometry,” *Ear & Hearing* 30(6): 635-652
- Liu, Y.-W., Sanford, C.A., Ellison, J.C., Fitzpatrick, D.F., Gorga, M.P., and Keefe, D.H. (2008). “Wideband absorbance tympanometry using pressure sweeps: System development and results on adults with normal hearing,” *J. Acoust. Soc. Amer*. **124**, 3708-3719.

MP3 vs. AAC

- MPEG-1 Layer III
  - Finalized in 1992
  - Known as MP3
  - Good at 128 kbits/s, per channel.

- MPEG-2/4 Advanced Audio Coding (AAC)
  - Finalized in 1997
  - Known as the Apple iTune music format.
  - Good at 128 kbits/s, per *stereo*.
  - Not compatible to MP3
    - Its old name: MPEG-2 NBC (Non-Backward Compatible).

Sound quality evaluation

- Judged by professional golden ears
- Played back in professional studio room
- Randomized, double-blinded R-A-B tests

Quality	Impairment
5	Imperceptible
4	Perceptible, but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

- Good compression algorithm scores > 4.0 for European Broadcast Union's "Sound Quality Assessment Materials".

Speech coding vs. audio coding

- Speech coding:
  - Parametric, based on production model
  - Sampling rate 16 kHz is usually sufficient
  - Targeting data rate <10 kbps
  - Not good for coding music

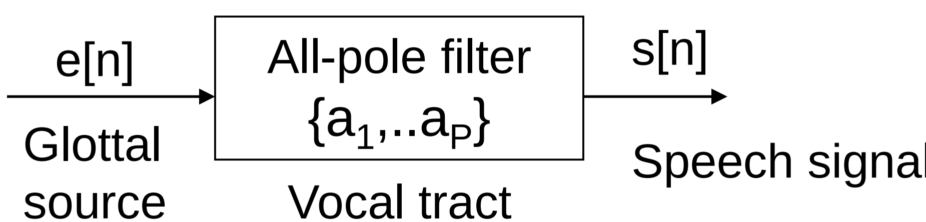
- Audio coding:
  - Non-parametric, based on perceptual model
  - Sampling rate 44.1 kHz or higher
  - Targeting data rate 128 kbps/stereo

DETOUR TO A SPEECH CODING STANDARD (CELP)

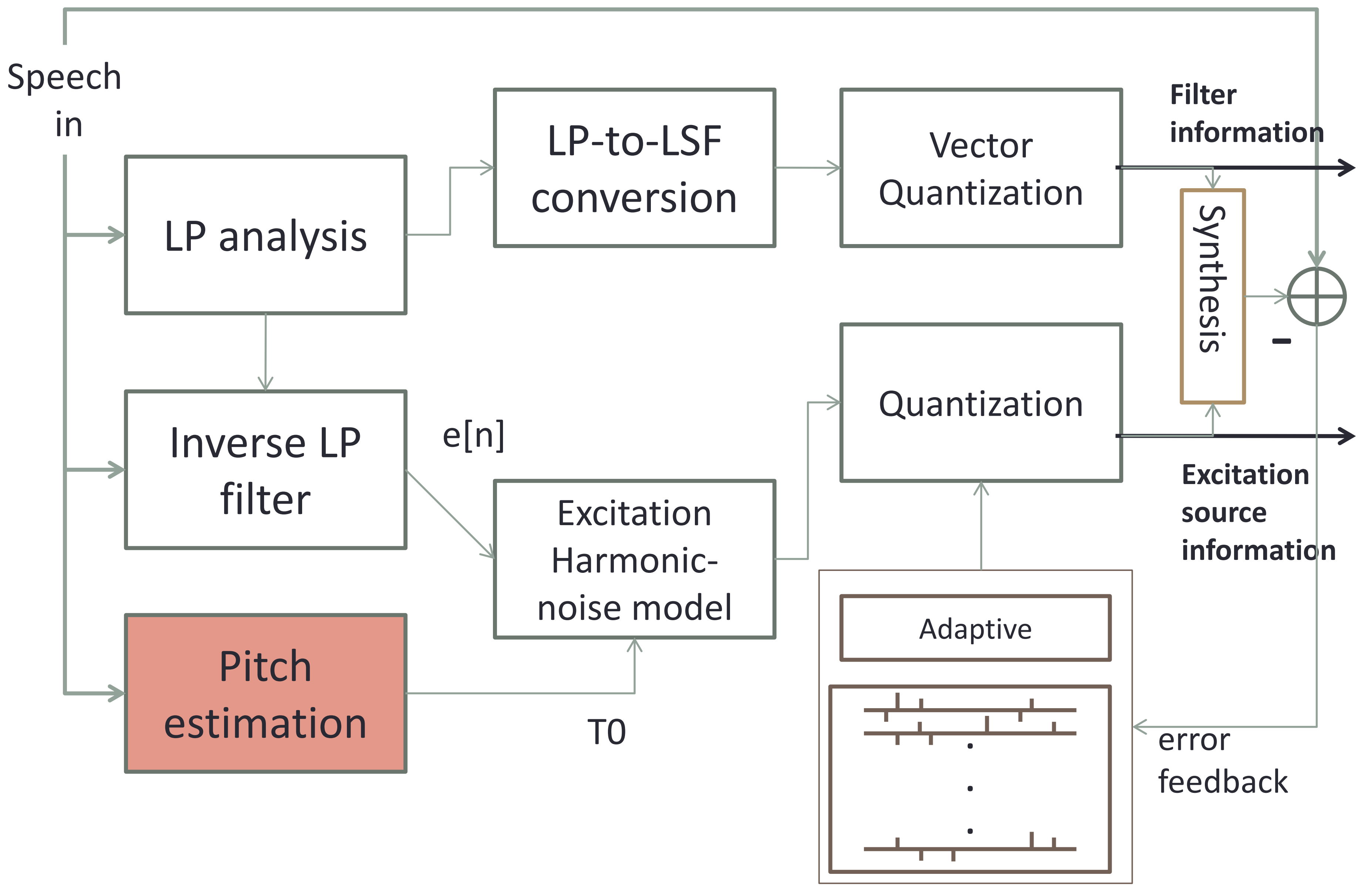
Possible applications of LP

- **Speech synthesis**: By replacing e[n] with a template, speech compression achieves <8k bits/s.
  - Codebook excited linear prediction (CELP)
  - key technology for voice over internet and wireless networks.

- **Speech recognition**: From $\{a_1,..a_P\}$, we can estimate
  - Vocal tract constriction
  - Frequency-envelope; formant structure.



Code-Excited Linear Prediction



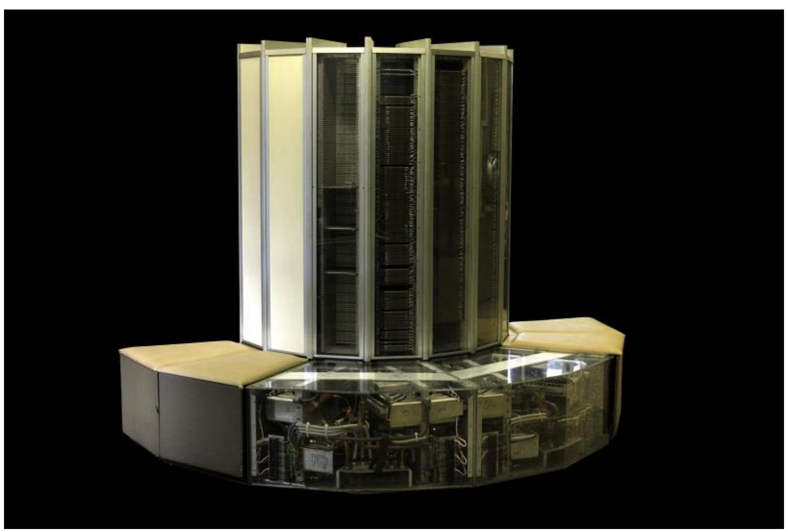
了不起的地方

- ¼ bit per sample x 8000 samples/sec = 2 kbps.
- Digital telephone land line = 64 kbps.
- 32:1 data compression, proposed in 1985!
  - At AT&T Bell Labs.
  - However…

Coding procedure was very expensive

"…It took 125 sec of **Cray-1** CPU time to process 1 sec of the speech signal."

- Cray-1 (1975): 80-MHz "super computer".



http://en.wikipedia.org/wiki/Cray-1

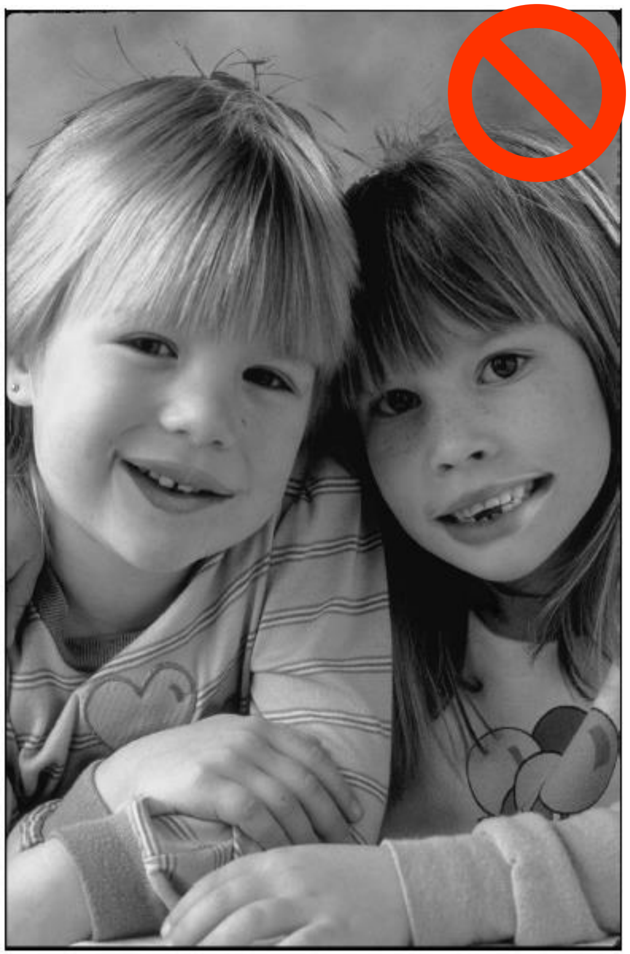
Topic II: Watermarking

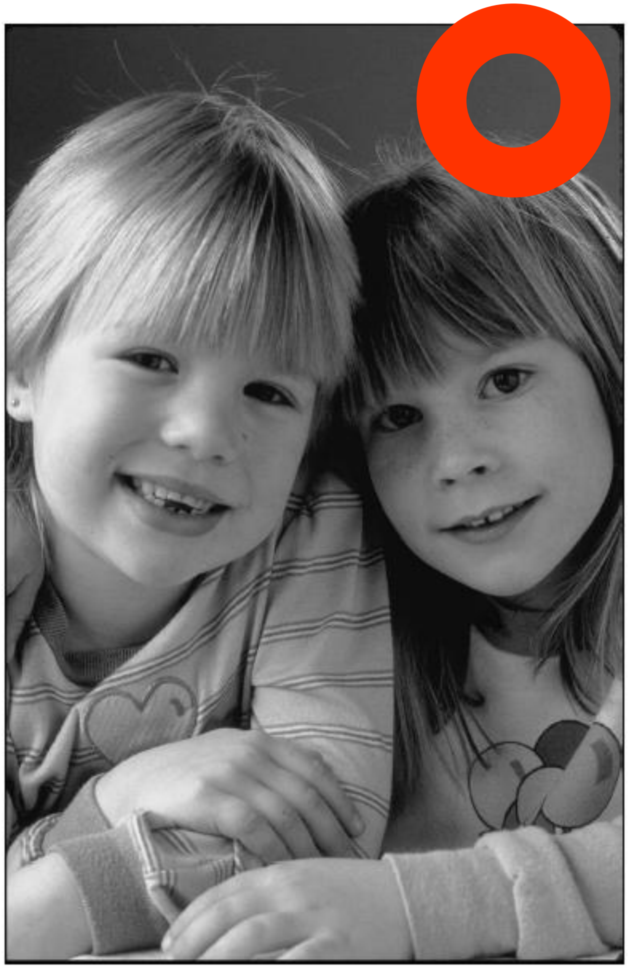
Bureau of Engraving and Printing,
United States Department of the Treasury
http://www.moneyfactory.com/

Digital watermarking

Adapted from: E. Lin et al. (2000).

“Detection of image alterations using semi-fragile watermarks”

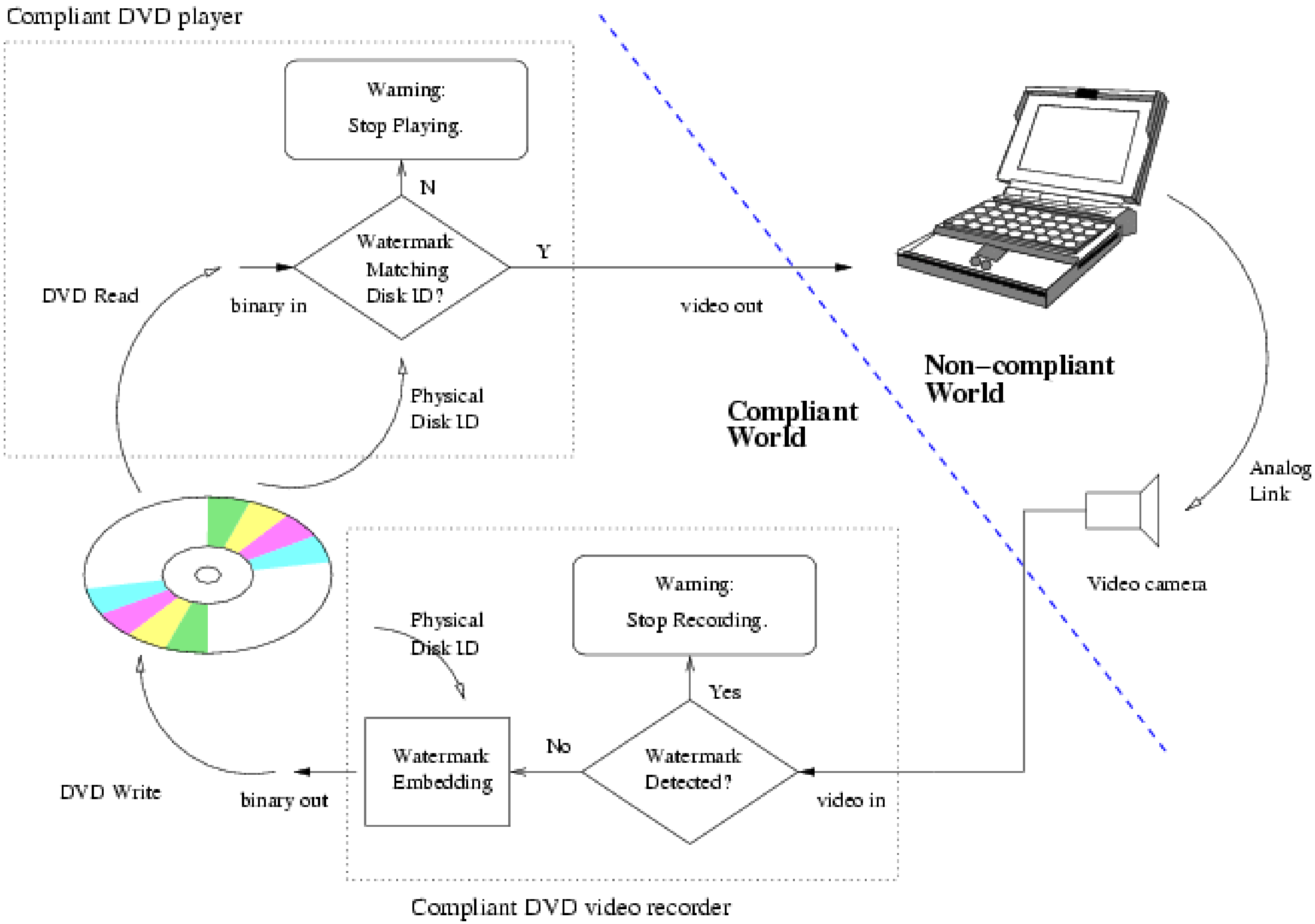




Application #1: forgery detection

Application #2: copy management

J.-P. Linnartz *et al*, *"System aspects of Copy Management for digital video"* (2000)



Application #3: automatic broadcast monitoring

J. Gertner, "Our Ratings, Ourselves", *the New York Times Magazine, April 10, 2005*

"Television and media will change more in the next 3 to 5 years than it's changed in the past 50."

~ Bob Luff, chief technology officer, Nielsen Media

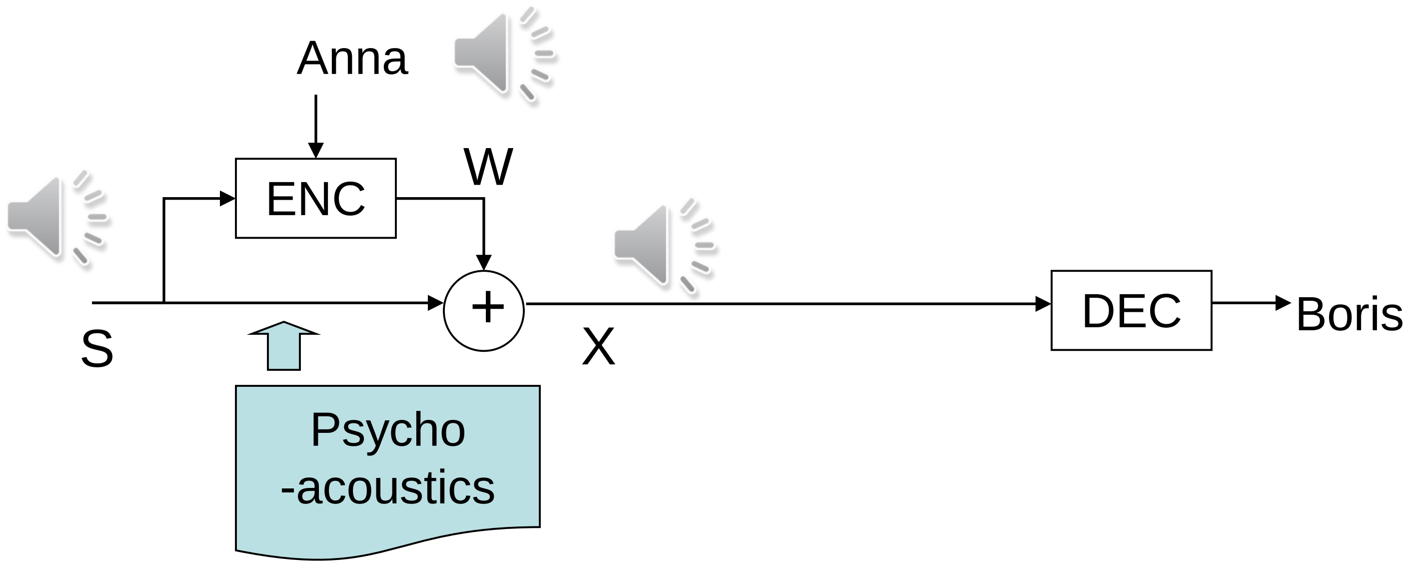
- TV on cellphone
- Radio on Web
- iPod, video-on-demand, PlayStation Portable…
  - It becomes difficult to measure people's exposure to media.

“Portable people meter” (PPM)

- Programs and commercials are embedded with acoustical watermarks
- A wearable device
  - Picks up the watermarks
  - Identifies programs
- System tested in Houston
  - By Arbitron Inc. (NYSE: ARB)

http://www.arbitron.com/portable_people_meters/home.htm

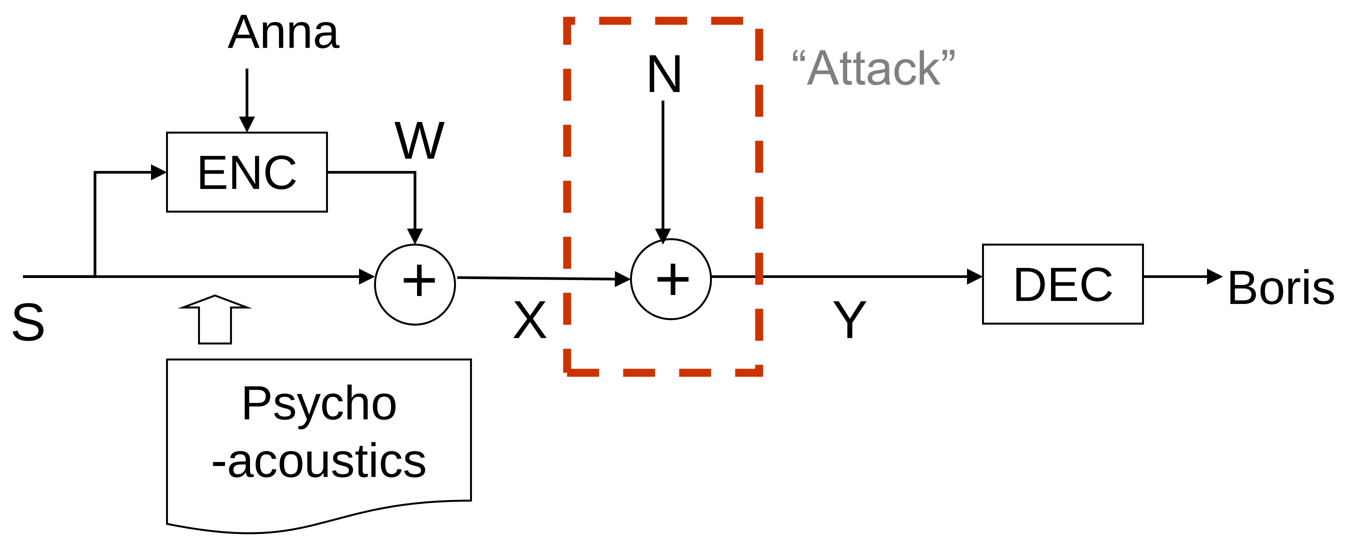
The principle of watermarking:
noise is signal and signal is noise.



W: Watermark
S: Signal

X: Watermarked signal

浮印植入者與破壞者的賽局
The game between embedder and attacker



W: Watermark
S: Signal

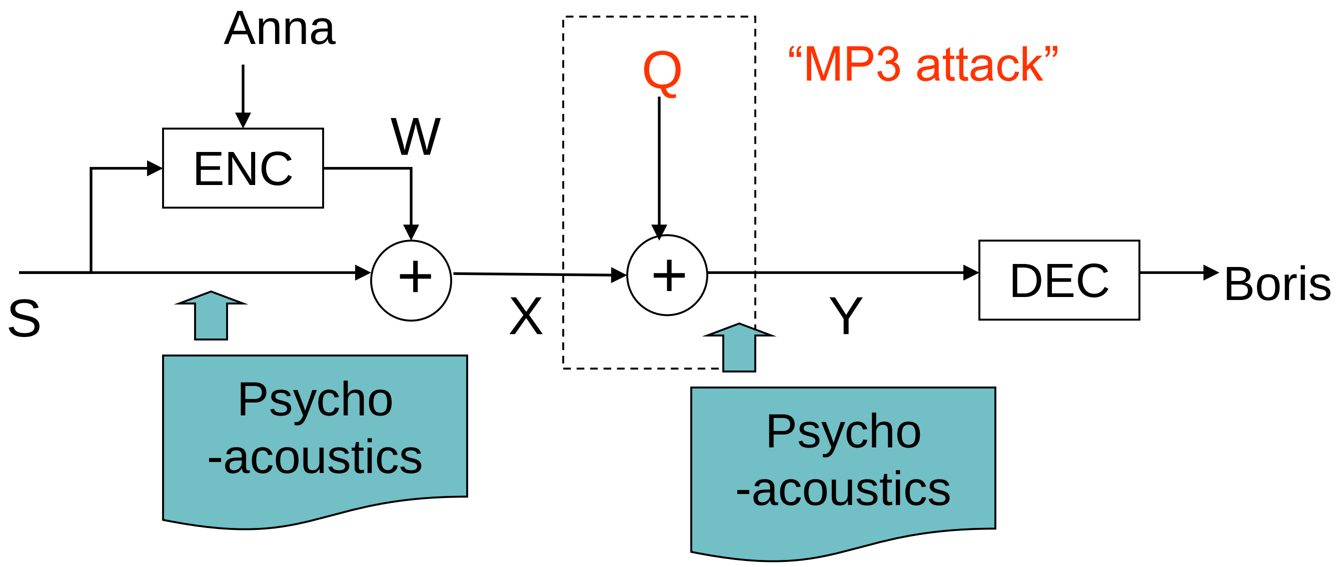
X: Watermarked signal
N: Noise
Y: Corrupted copy of X

Common misconception

- audio watermarks are *useless* because
  - They are inaudible
  - MP3 removes everything inaudible
  - therefore, MP3 erases watermarks

Rules of the Game

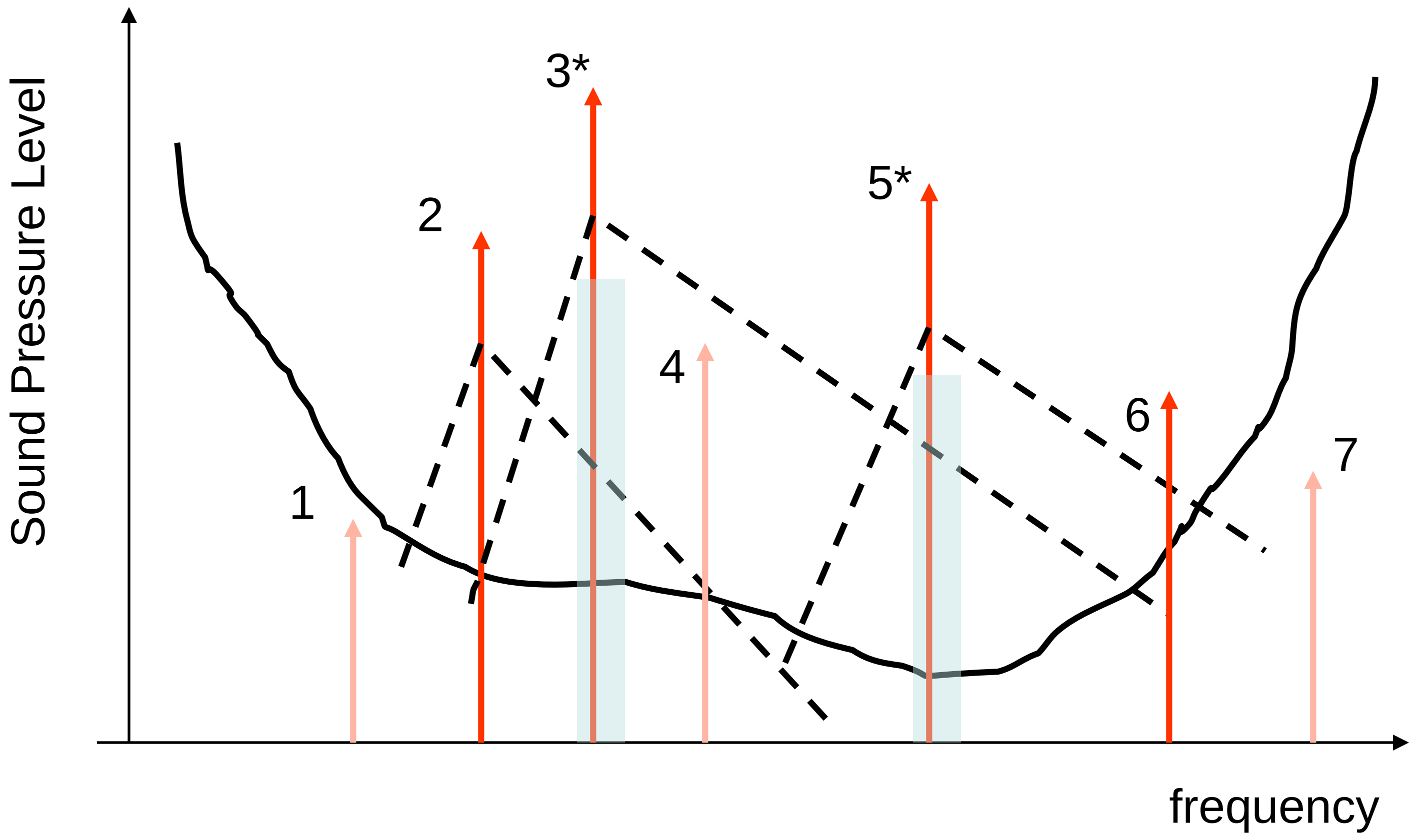
1. Embedder plays first.
2. Attacker also needs to consider psychoacoustics.



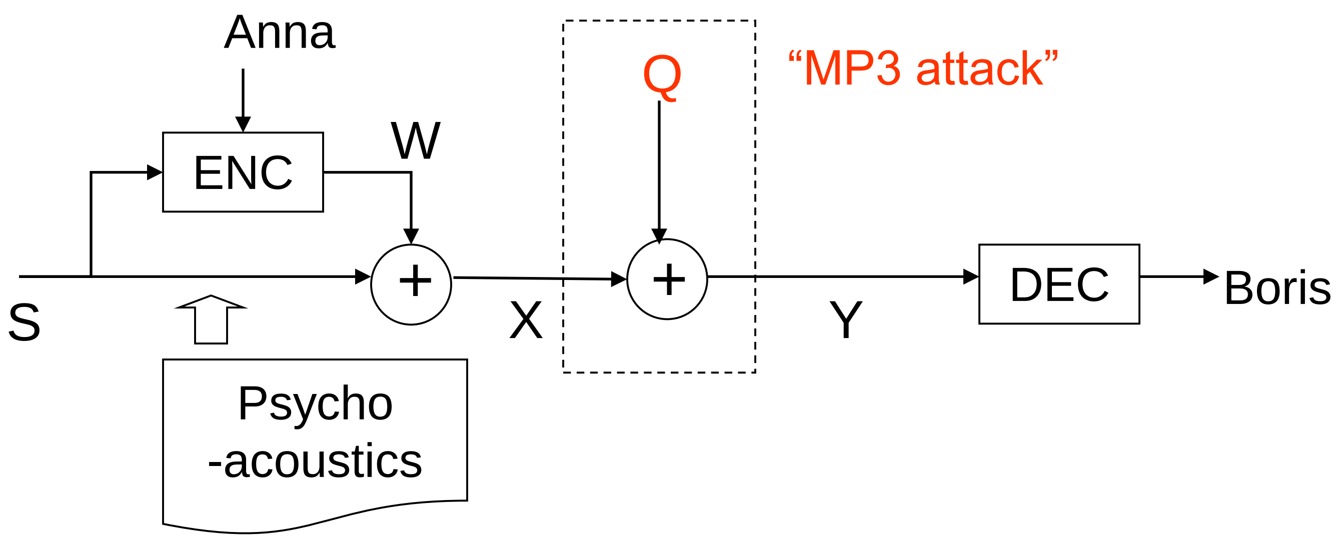
Optimal embedding strategy:

Watermark is indistinguishable from signal,

but decodable by *coherent detection* via a private key.



Quantization introduces noise; Watermark-to-noise ratio is improved by averaging.



W: Watermark
S: Signal
Q: Quantization noise

Y = S + W + Q.

Duality of data compression and watermarking

	Compression	**Watermarking**
Noise to be hidden	Quantization noise	Watermarks
Information theoretic models	Source coding (Rate-distortion problem)	Channel coding (communication)
Rules	• Encoder: secret • Decoder: public	• Encoding: public* • Decoder: blind

*** except for the crypto-key.**

Usefulness of audio watermarking is still uncertain in future

- Apple relinquished **digital right management** (DRM) for iTune since Feb. 2009
  - DRM = Digital Restriction Management?
  - It does not help business.

- Microsoft licensed audio watermarking technology to Active Content Corp. in Aug. 2007
  - “Personalized” ads bundled to music

Summary

- Audio Data Compression
  - Psychoacoustic masking enables 10:1 compression in AAC.
  - 6 dB SNR improvement for every extra bit.
  - Format is standard, but encoding is commercial secret.

- Audio Watermarking
  - Using psychoacoustics for data hiding
  - Robust to MP3
  - Future application is uncertain.