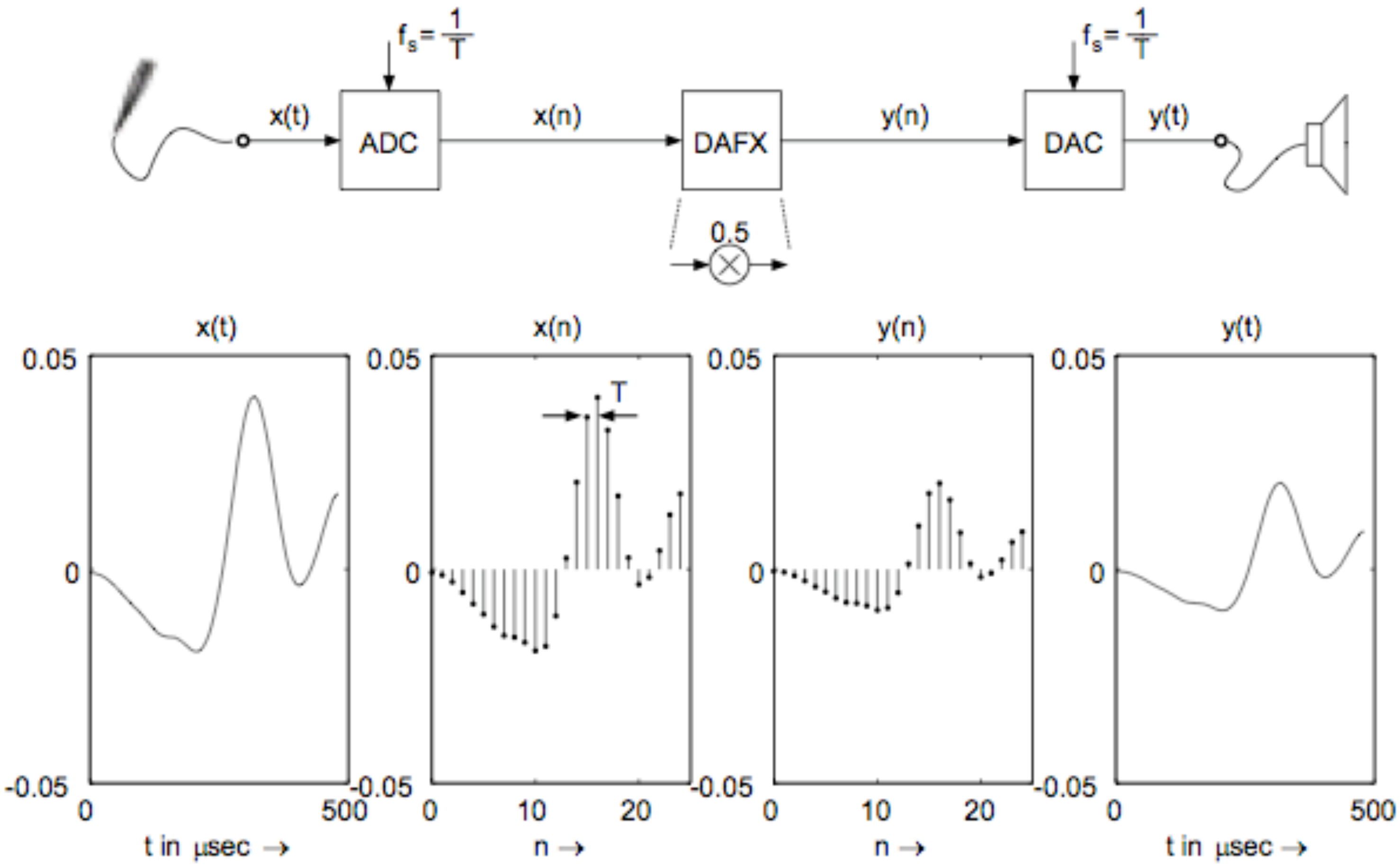
Digital audio processing revisited

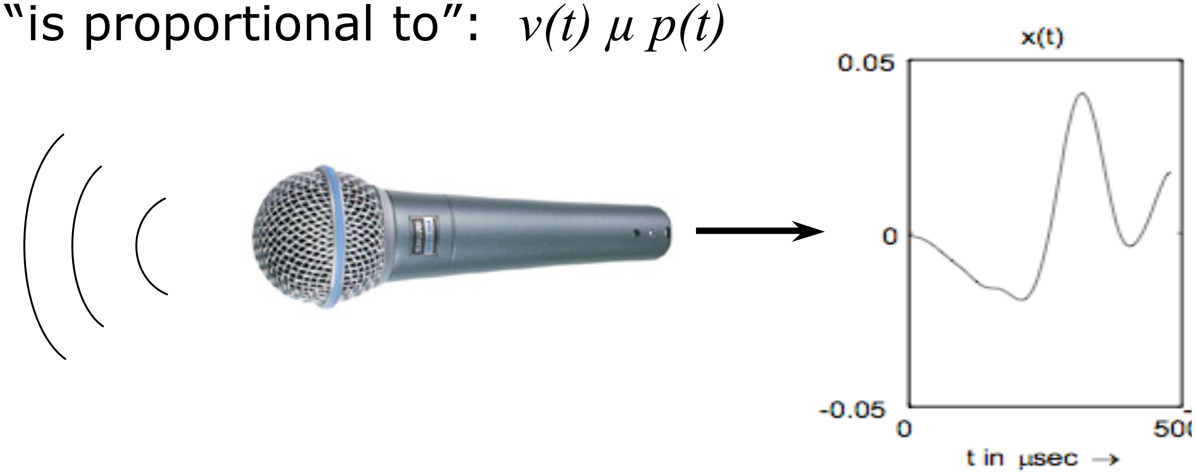
Juan P Bello

Digital audio processing



Microphones

- Sound is an energy disturbance that propagates through a medium as a wave
- Commonly, the medium is air, thus the sound wave produces variations of air pressure
- A microphone is a transducer (i.e. a device that converts energy or information from one form to another).
- Specifically, the microphone converts air pressure into voltage levels, thus generating an electrical signal *analogous* to the mechanical one.
- The following expression notates the relationship between voltage and pressure in a microphone, where the symbol µ means "is proportional to": $v(t)\ \mu\ p(t)$



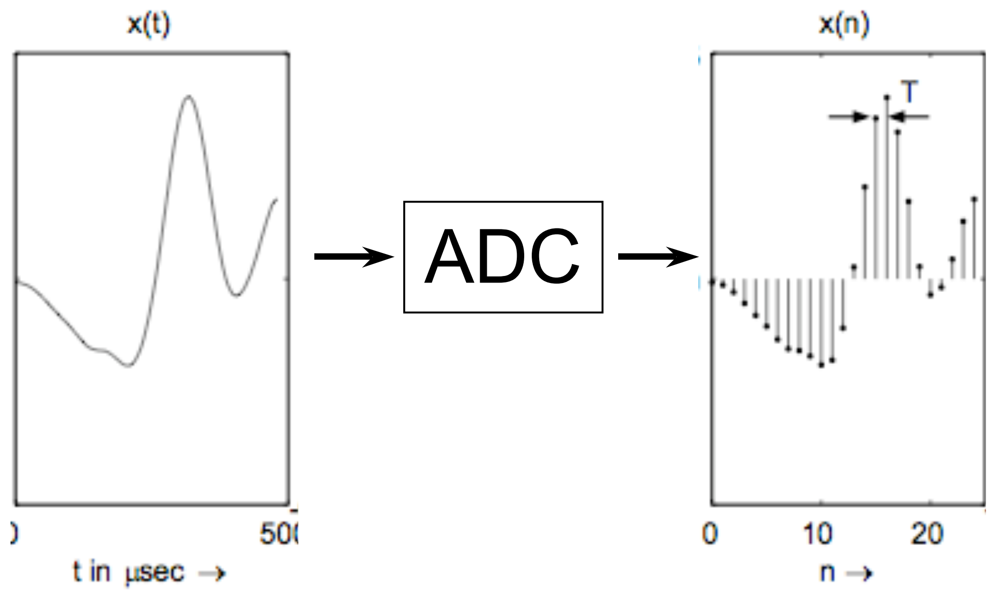
ADC

- The conversion of an analog (continuous) signal x(t) into a discrete sequence of numbers x(n) is performed by an Analog-to-digital Converter (ADC)
- The ADC samples the amplitude of the analog signal at regular intervals in time, and encodes (quantizes) those values as binary numbers.
- The regular time intervals are known as the sampling period (T_s) and are determined by the ADC clock.
- This period defines the frequency at which the sampling will be done, such that the sampling frequency (in Hertz) is:

$$fs = \frac{1}{T_s}$$

- The accuracy of the quantization depends on the number of bits used to encode each amplitude value from the analog signal.

ADC



- The outgoing sequence x(n) is a discrete-time signal with quantized amplitude
- Each element of the sequence is referred to as a sample.

$$\ldots, x[n-1], x[n], x[n+1], \ldots$$

Discrete signals

- An example discrete signal is a real sinusoid, which can be described as:

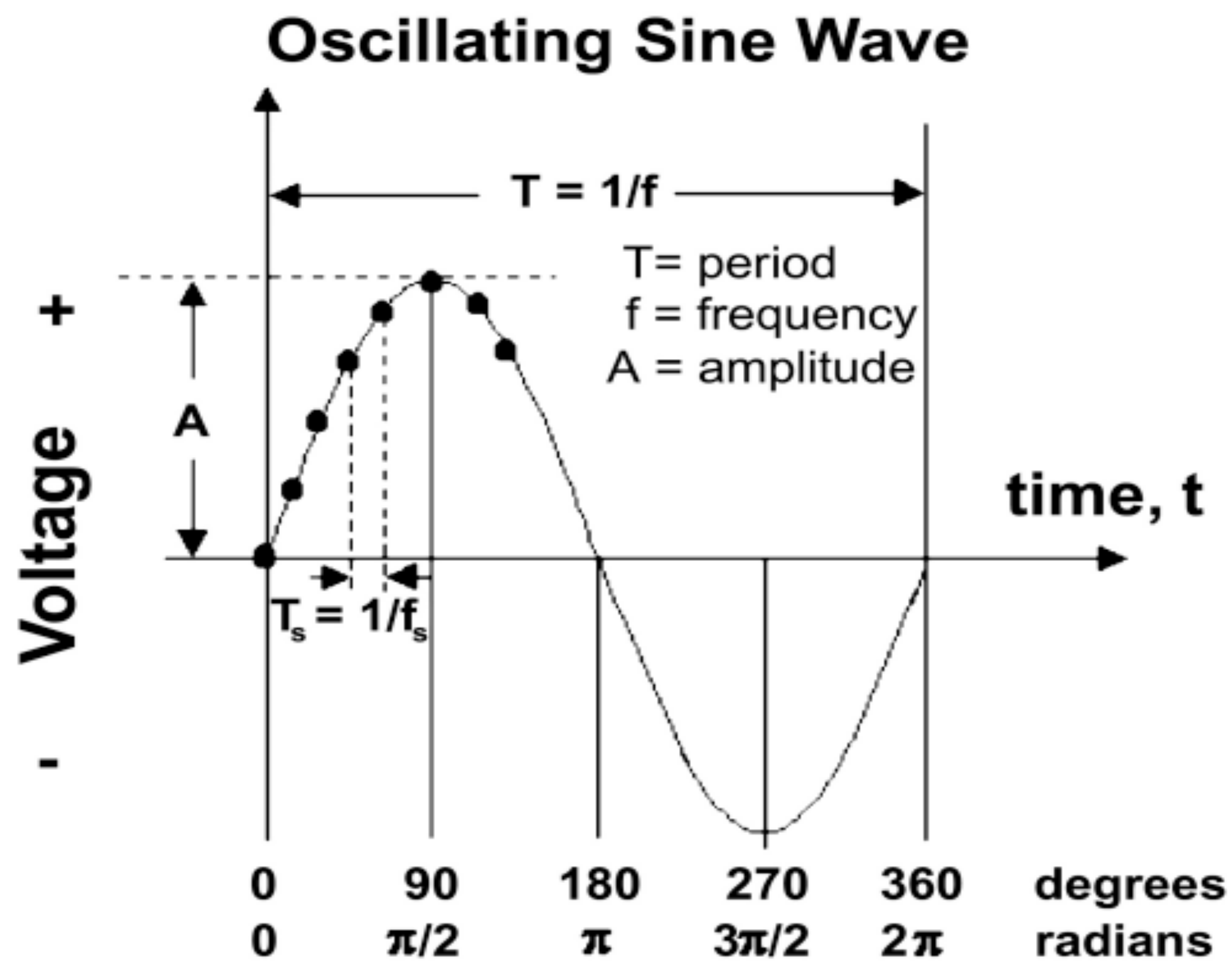
$$x[n] = a\cos(\omega n + \phi)$$

- where a is the amplitude, ω the angular frequency, and ϕ the initial phase. At sample number n, the phase is equal to $\phi + \omega n$.
- A sinusoid is an example of simple harmonic motion.
- Because each cycle is completed in a constant amount of time, the motion of the wave is *periodic*, i.e. there is a T > 0 that satisfies the equation:

$$f(n) = f(n + \mathrm{T}), 0 < \mathrm{T} < \infty$$

- The number of cycles completed per second is the frequency of the wave, and the inverse of the frequency is its period.

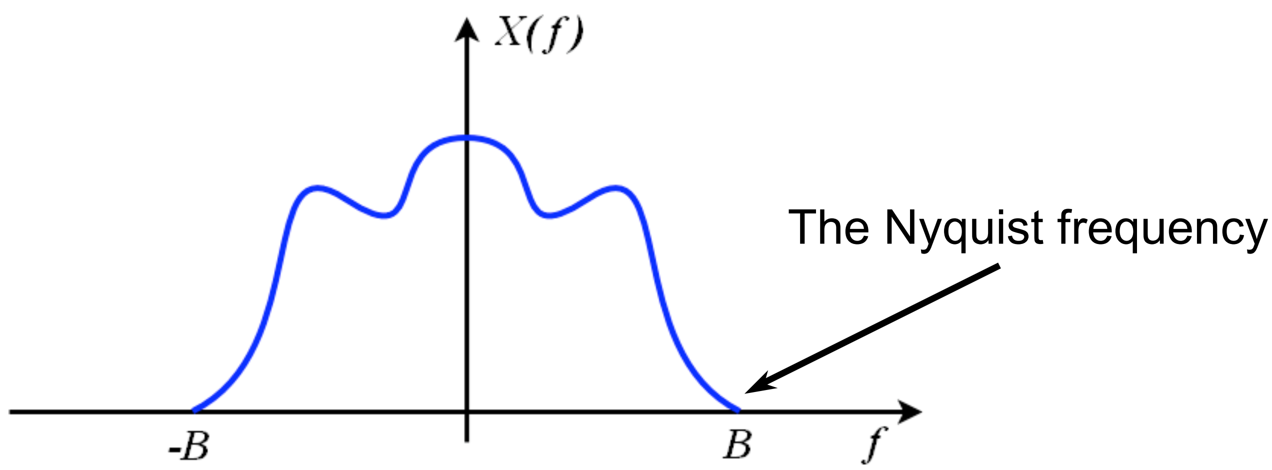
Discrete signals



- A sine and a cosine are differentiated only by a phase difference of a quarter cycle (π/2)

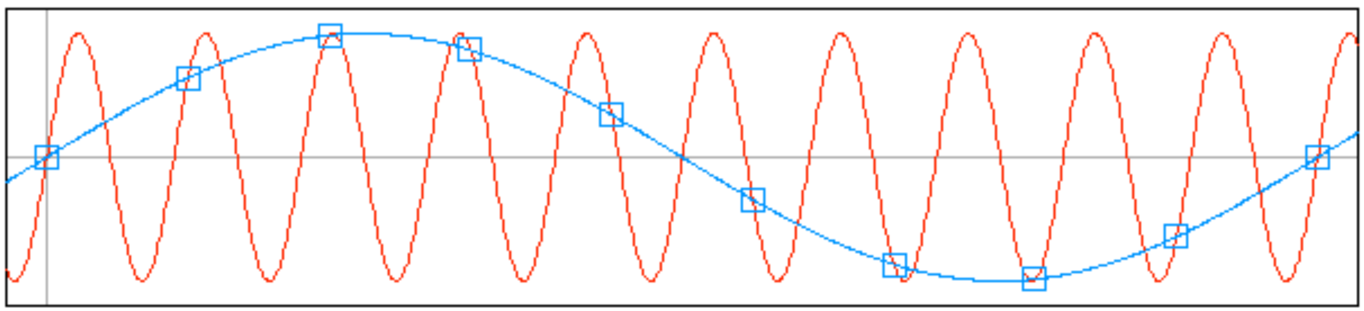
The sampling theorem

- Sampling is the process of converting a continuous signal into a discrete sequence
- Our intuition tells us that we will loose information in the process
- However this is not necessarily the case and the sampling theorem simply formalizes this fact
- It states that "in order to be able to reconstruct a ***bandlimited*** signal, the sampling frequency must be at least ***twice*** the highest frequency of the signal being sampled" (Nyquist, 1928)

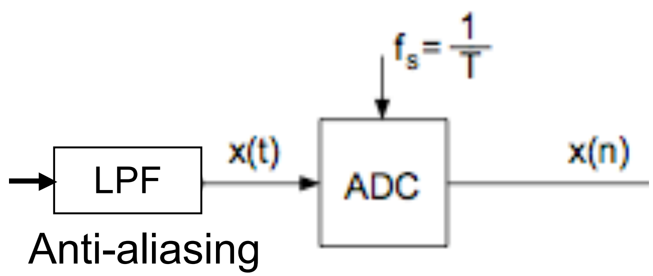


Aliasing

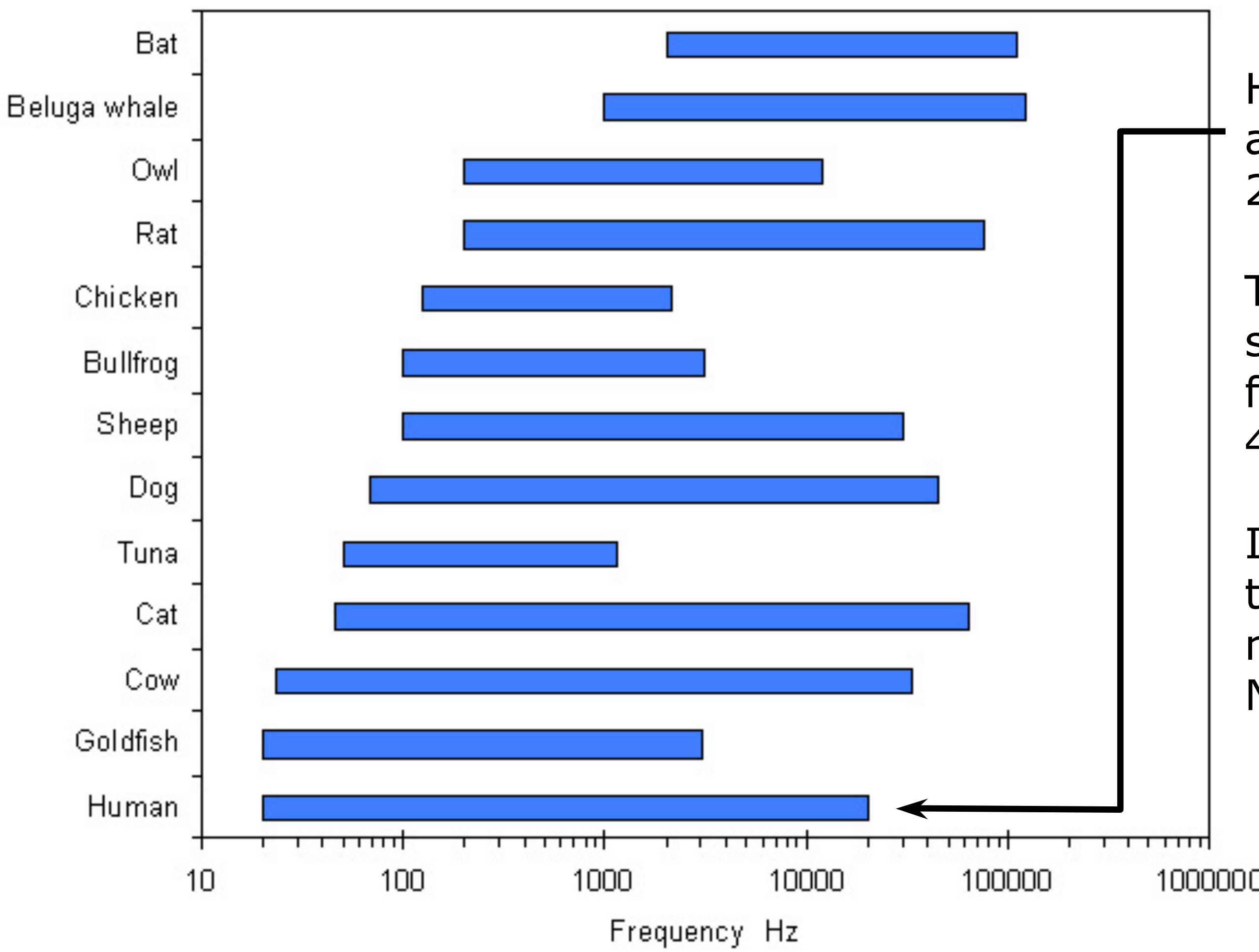
- What happens when $fs < 2B$
- There is another, lower-frequency, signal that share samples with the original signal (an alias).



- Related to the wagon-wheel effect: http://www.michaelbach.de/ot/mot_strob/index.html



Hearing frequency range

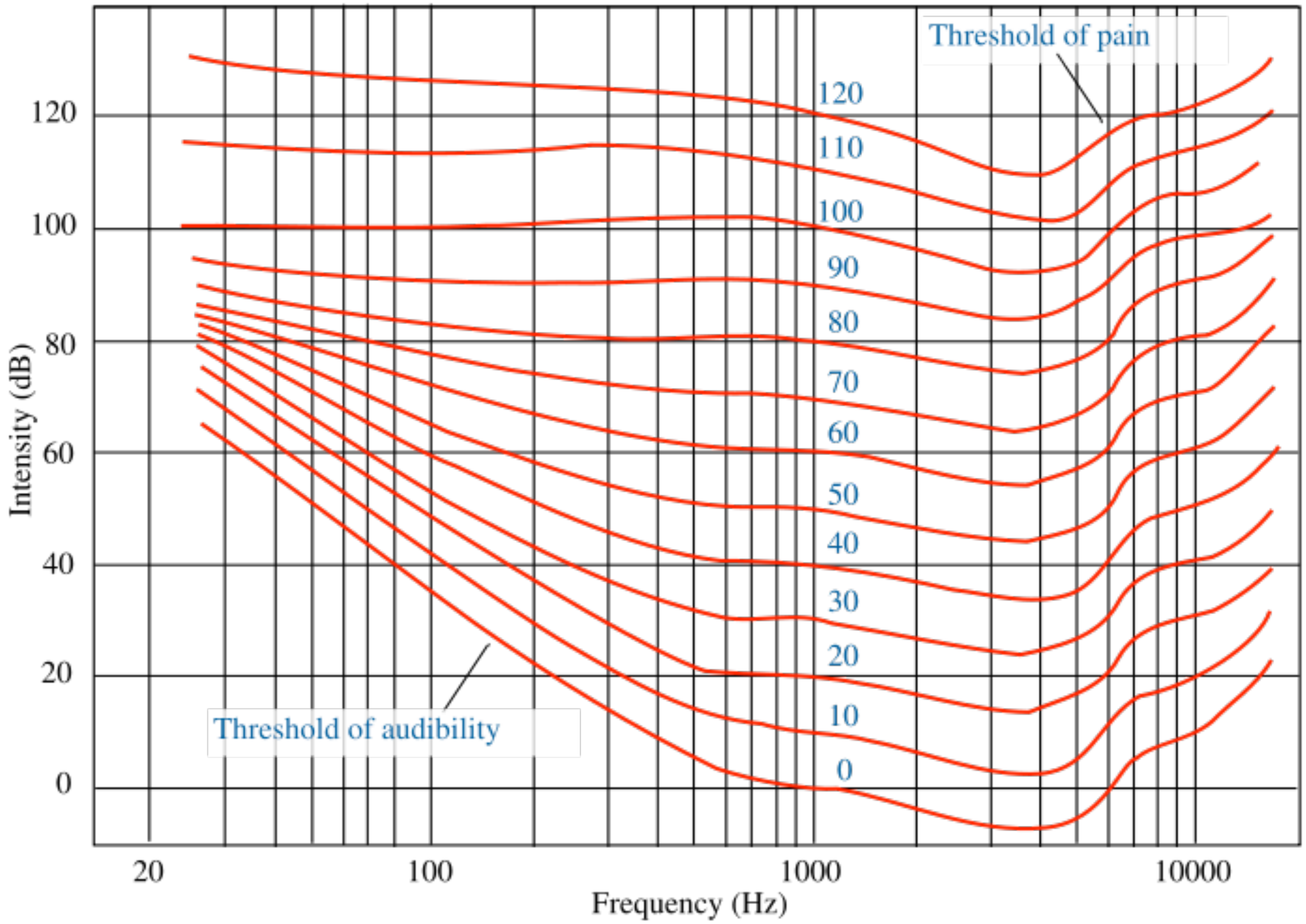


Human hearing is widely accepted to lie in the 20-20kHz range

Thus main reason for standard sampling frequencies to be of 44.1kHz and 48kHz

In digital synthesis we then have to be careful not to exceed the Nyquist frequency

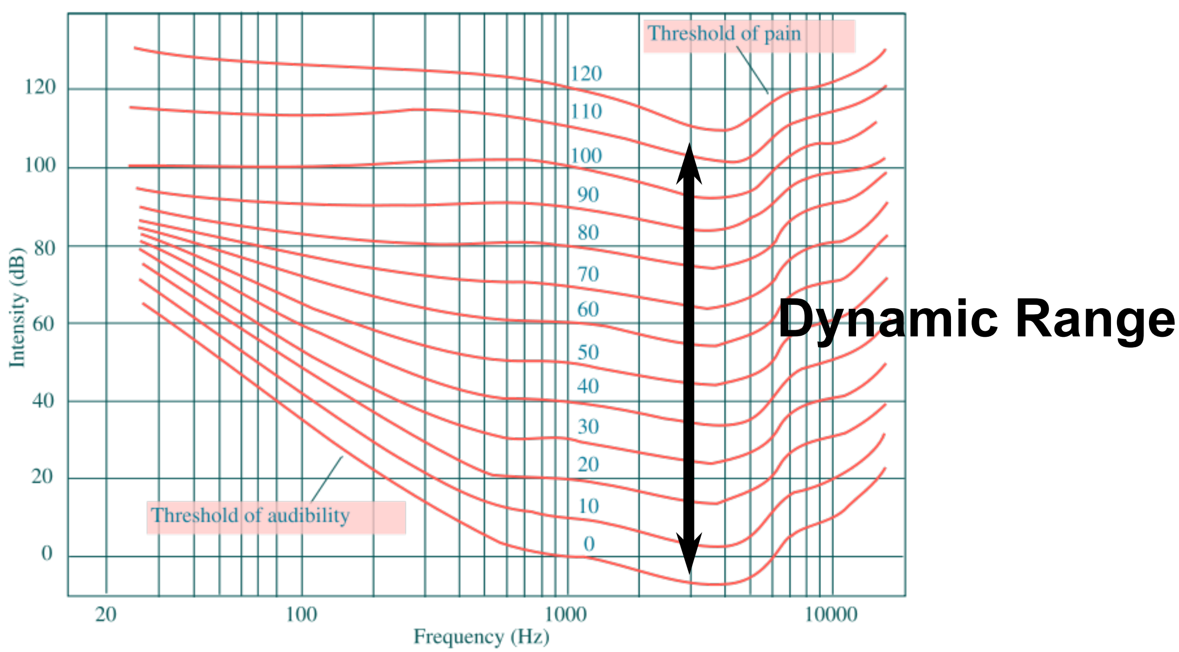
Loudness



- dB = 10 * $\log_{10}$(level/reference level) - Levels of intensity or power
- Reference level = 0dB = 10^{-12} watts per square meter (threshold of hearing)

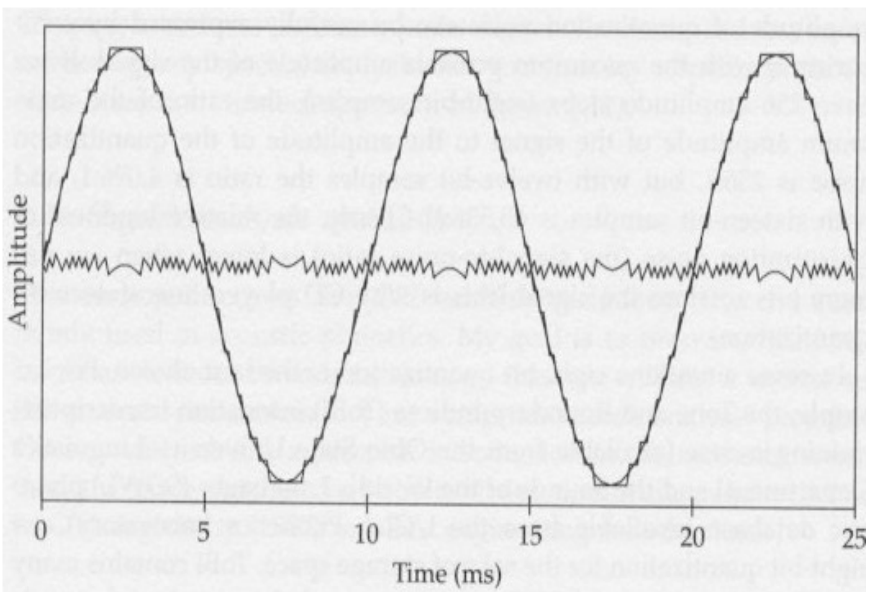
Dynamic Range

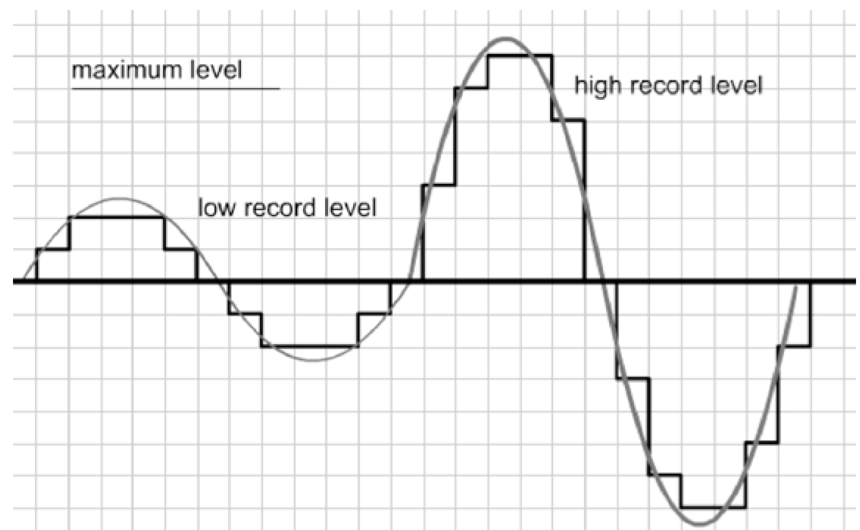
- Threshold of hearing is ~0dB and threshold of pain is ~125dB
- Dynamic range of a system: difference between the loudest and softest sound that a system can produce (measured in dB)
- On a linearly encoded PCM streams it is roughly: # of bits * 6



Quantization noise

- Is the distortion produced by the rounding-up of real signal amplitude values during the ADC process to the values “allowed” by the bit-resolution of each sample.
- The difference in level between the intended signal and the noise arising from quantization is the signal-to-quantization-noise ratio (SQNR)
- This depends on the quantization accuracy (# of bits) and the signal itself.





- Example: a sound with progressively worsening quantization noise:

Low-level quantization noise

- Sounds just above silence are degraded most severely by the quantization noise, because all of the variation is captured by the least significant bit.

- This is known as low-level QN, i.e. a square wave produced by 1-bit variations triggered when the signal has a very low amplitude.
- This noise can be critical as square waves are rich in odd harmonics, that can even extend beyond the Nyquist frequency producing aliasing.

- Solutions to this problem include:

1. Increasing the bit resolution (the level of noise is “inversely proportional” to the number of bits per sample)
2. Adding dither, i.e. low-energy analog noise added prior to the AD conversion, hence randomizing the quantization noise. Low-level uncorrelated wide-band noise (amplitude typically LSB/2) is less intrusive than square wave noise.

Dithering

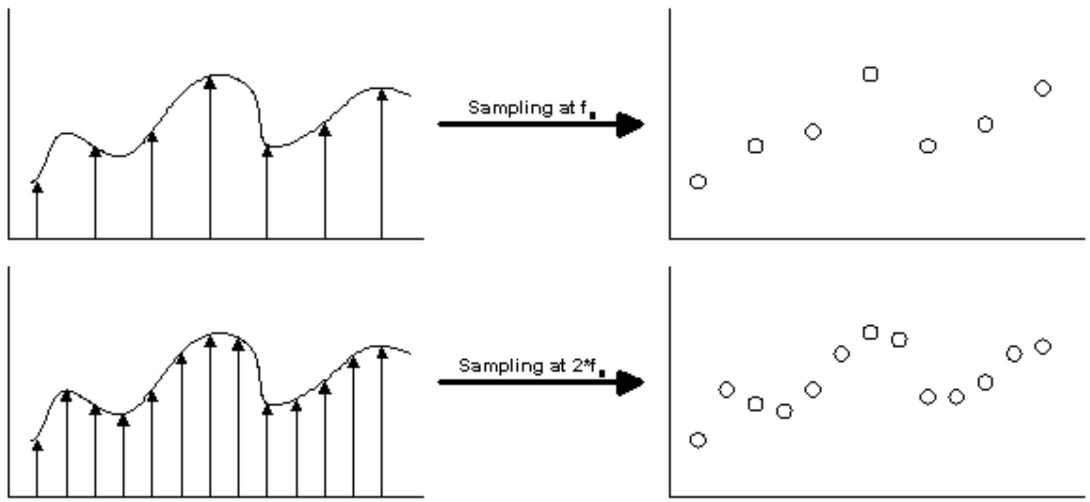
Original

8-colors no dither

8-colors + dither

Oversampling

- If the desired sampling rate is X, *oversampling* will perform the analog-to-digital conversion at some faster rate, such as 2X.



- The technique can be used to: minimize aliasing, noise reduction and increase accuracy beyond that provided by the wordlength.
- It widens the range of the frequency spectrum thus reducing the (uniformly distributed) noise below the Nyquist frequency.
- When the final filtering is performed, the residual quantization noise in the audible signal will be less: 4X oversampling yields a 6 dB reduction (12 dB for 8X oversampling)

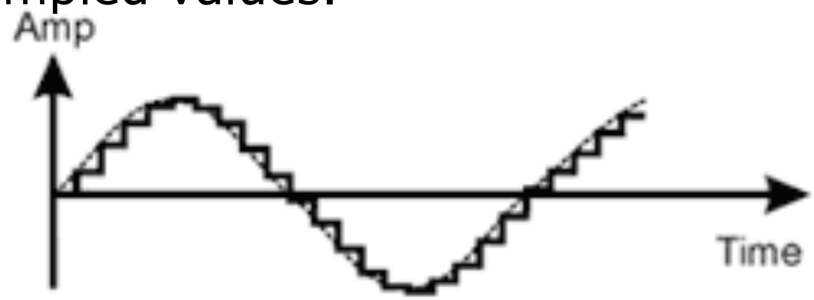
Storage Requirements

Type	Wordlength	SamplingRate	SQNR	Bytes/minute/channel
CD	16 bits	44100	96 dB	5,292,000
CD	16 bits	48000	96 dB	5,760,000
DVD	24 bits	88200	144 dB	10,584,000
DVD	24 bits	96000	144 dB	11,520,000
DVD	24 bits	192000	144 dB	23,040,000

Storage requirement = f_s * *wordlength* * *duration* * *channels*

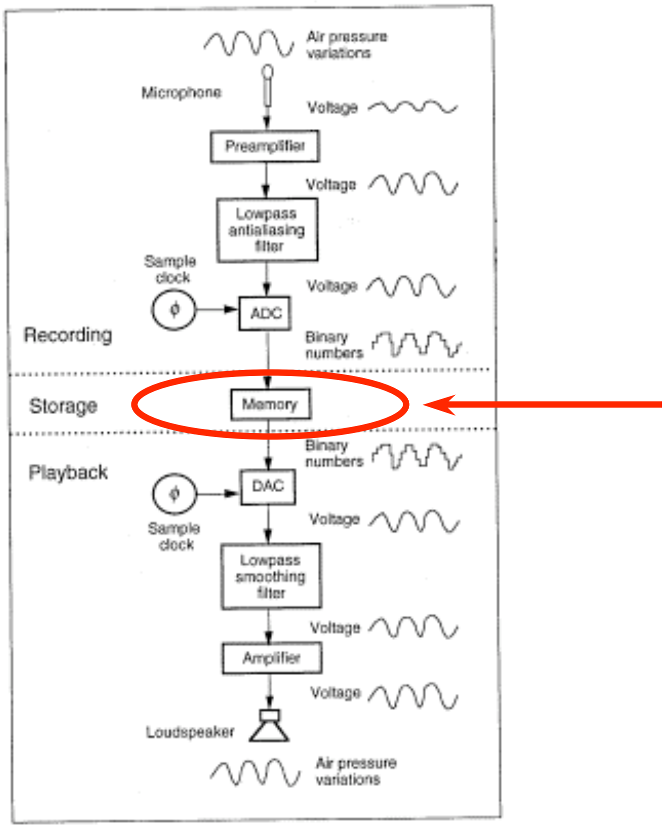
DAC and Imaging

- Just as we used an ADC to go from x(t) to x(n), we can turn a discrete sequence into a continuous voltage-level signal using a Digital-to-analog converter (DAC).
- However, the quantized nature of the digital signal produces a "Zero-Order Hold" effect that distorts the converted signal, introducing some step (fast) changes.
- This distortion is know as *imaging*.
- To avoid this, we use a low-pass filter after the DAC, such that it smoothes out those fast changes.
- The filter, known as an anti-imaging filter (AKA smoothing or reconstruction filter), discards signal components above the Nyquist frequency, thus performing a simple interpolation between the sampled values.



High Sampling Rate

Digital Recording and Playback



This is not only storage, this is our digital system!

That system is supposed to process the signal somehow

Still we do not know anything about our system

Digital systems

- The digital system can be seen as an algorithm that operates on the discrete input sequence x(n)
- The output of such a system is the sequence y(n)
- The simplest of such systems are known as Linear Time-invariant (LTI) systems
- As the name indicates they must be time-invariant: i.e. their behavior does not change over time; and linear: they fulfill the following condition:

$$\begin{aligned} &if \quad x(n) = A \cdot x_1(n) + B \cdot x_2(n) \\ &then \quad y(n) = A \cdot y_1(n) + B \cdot y_2(n) \end{aligned}$$

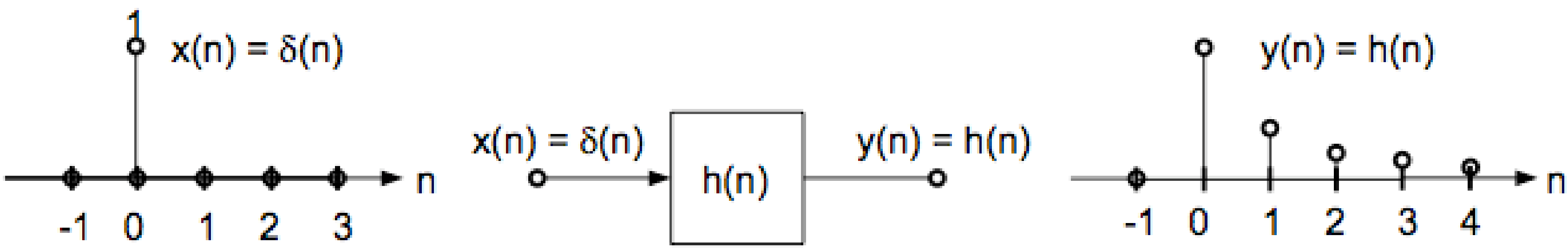
- For any constant A and B, and for a system where $y_i(n)$ is the output of $x_i(n)$, thus satisfying the superposition and scaling properties

Impulse response

- The input/output relations on a LTI system can be characterized using a test signal
- A commonly-used test signal is the unit impulse, defined as:

$$\delta(n) = \begin{cases} 1 & n = 0 \\ 0 & elsewhere \end{cases}$$

- If we apply a unit impulse to a digital system we obtain $y(n) = h(n)$, the impulse response of the system.
- A digital system can be completely characterized by its impulse response

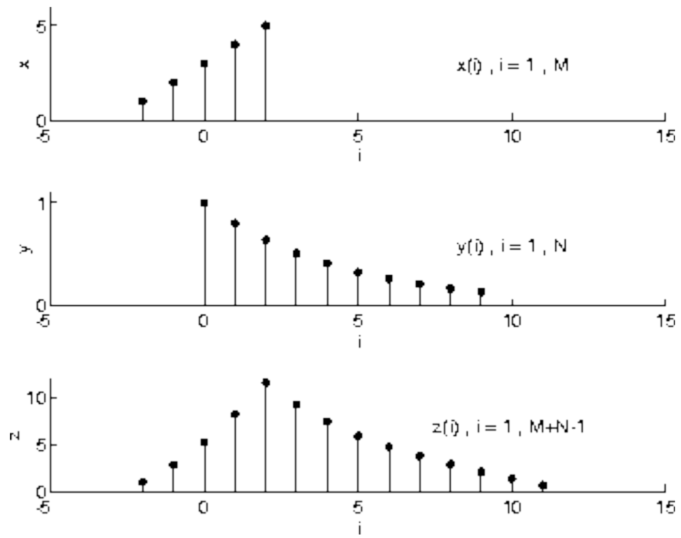


Discrete convolution

- Since we know the impulse response h(n) of a given system, we can calculate its response to ANY input signal x(n) by convolving the input with its impulse response:

$$y(n) = x(n) * h(n) = \sum_{m=-\infty}^{m=\infty} x(n) \cdot h(n-m)$$

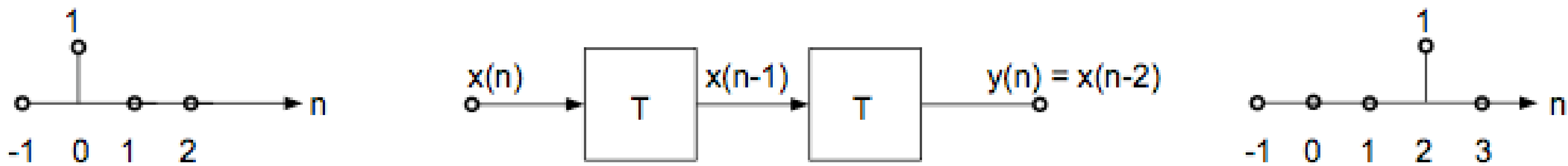
- A convolution represents the amount of overlap between x(n) and a reversed and temporally-shifted version of h(n)



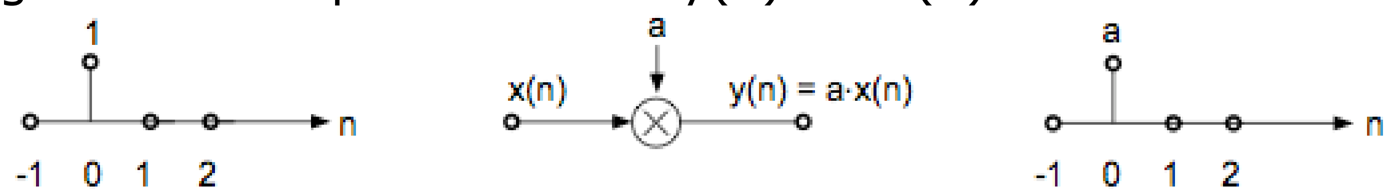
http://mathworld.wolfram.com/Convolution.html

Basic systems

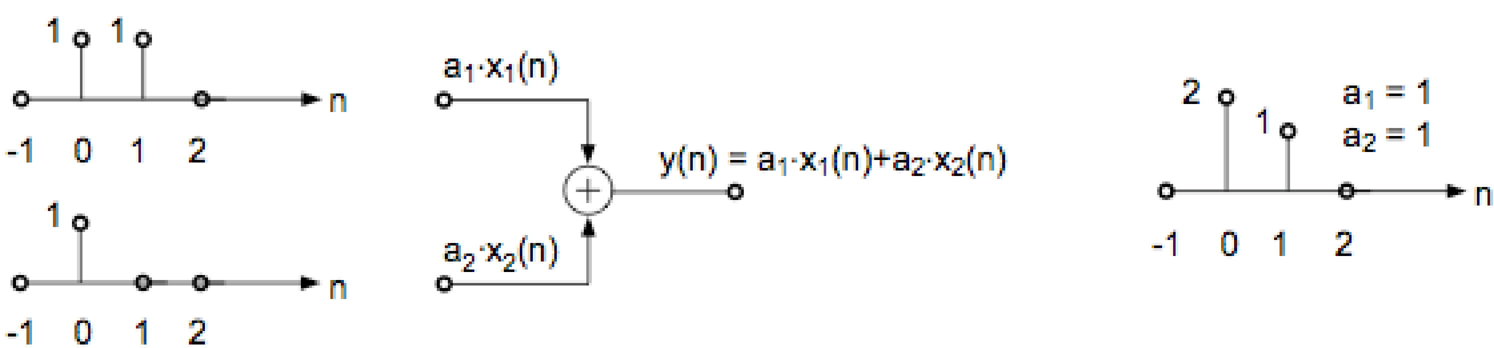
- A 2-sample delay can be described by the relation: $y(n) = x(n-2)$



- A gain of a is represented as: $y(n) = ax(n)$



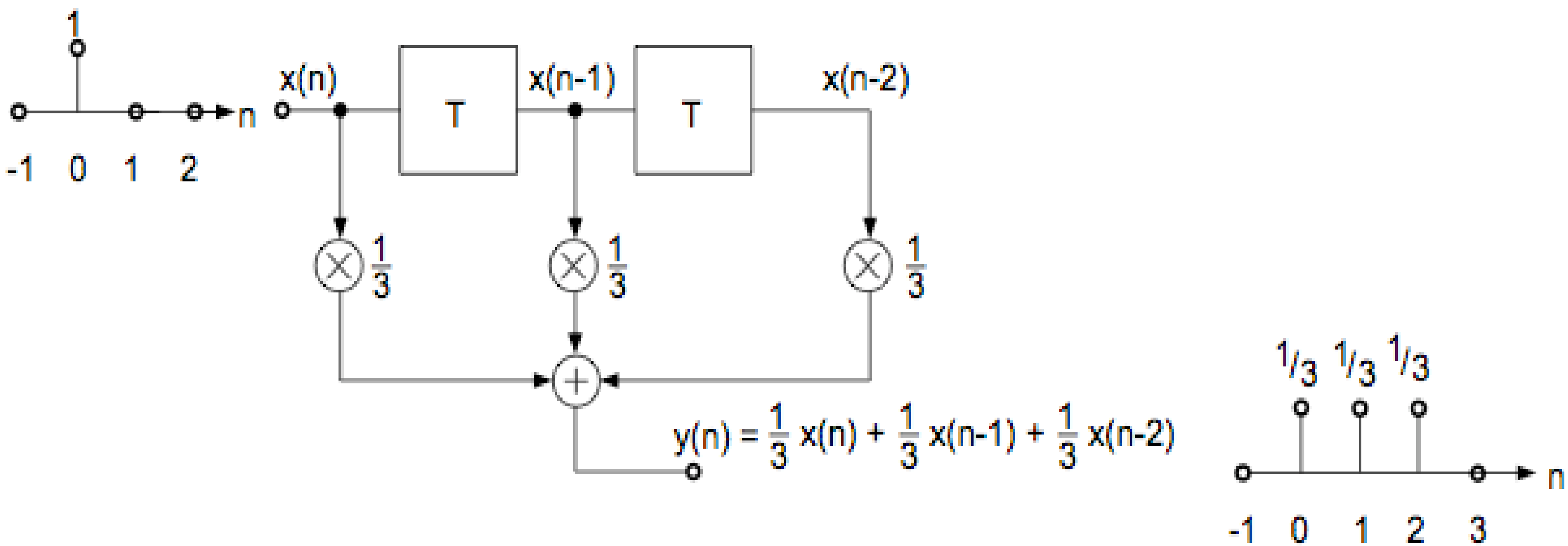
- The addition (mixing) of two inputs is: $y(n) = a_1x_2(n)+a_2x_2(n)$



Basic systems

- By combining the previous systems we can obtain a typical digital system:

$$y(n) = \frac{1}{3}x(n) + \frac{1}{3}x(n-1) + \frac{1}{3}x(n-2)$$



Transfer function

- However, the temporal relations between input and output are not all we can use to describe the system
- The frequency-domain behavior of a digital system specifies which input frequencies will be passed, rejected or emphasized.
- This behavior can be described using the transfer function H(z) and the frequency response H(f) (that will be discussed later)
- The transfer function is obtained by calculating the Z-transform:

$$X(z) = \sum_{n=-\infty}^{\infty} x(n) \cdot z^{-n}$$

- Of the impulse response h(n) as:

$$H(z) = \sum_{n=-\infty}^{\infty} h(n) \cdot z^{-n}$$

Causality and stability

- Some common Z-transforms:

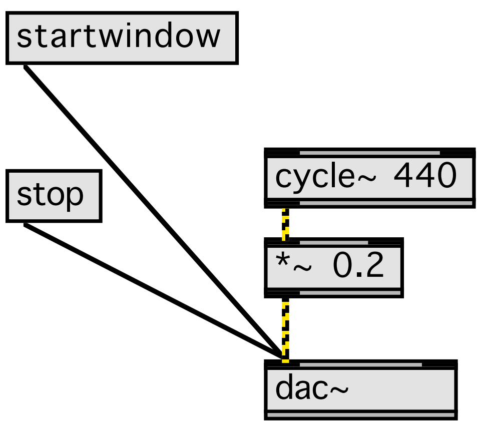
$x(n)$	$X(z)$
$x(n-M)$	$z^{-M} \cdot X(z)$
$\delta(n)$	1
$\delta(n-M)$	z^{-M}

- Finally, to be realizable, digital systems must be:

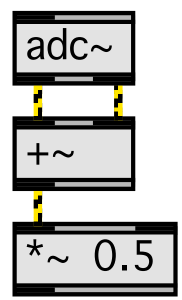
1. Causal: the system cannot react to an input before it is received
2. Stable: the sum of the absolute values of h(n) has to be less than infinite

Basic Systems in MSP

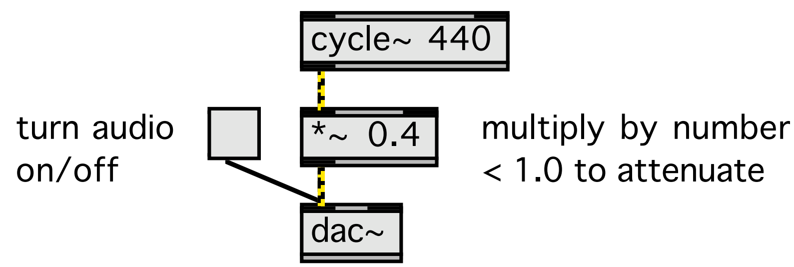
- MSP is a set of extensions to Max that provide for audio analysis, processing and synthesis
- All MSP objects end with a tilde '~' to indicate audio-rate processing. This because the tilde vaguely resembles a sine wave.



Send any discrete sequence to the DAC

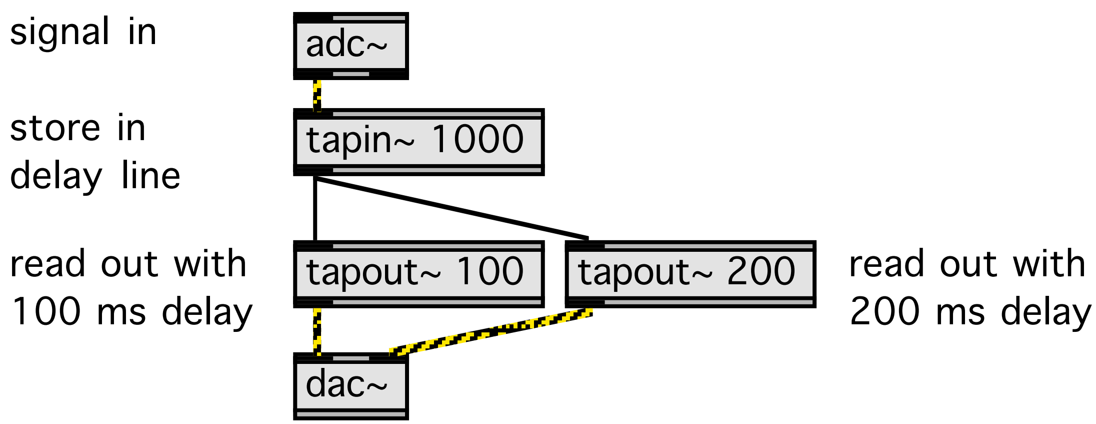


Mix



Change gain

Basic systems in MSP



- A tapin~ object saves some amount of its input signal in a buffer whose size is specified by the object's argument (here 1000 milliseconds).
- Any tapout~ objects connected to the outlet of a tapin~ share that same buffer, reading samples out after a delay.

Useful References

- Zölzer, U. (Ed). "DAFX: Digital Audio Effects". John Wiley and Sons (2002)
  - Chapter 1: Zölzer, U. "Introduction".
- Pohlmann, K. "Principles of Digital Audio". McGraw-Hill, Inc. (1995)
- Roads, C. "The Computer Music Tutorial". MIT Press (1996)