Aalto University
School of Electrical Engineering

Communication acoustics
Ch 3: Signal processing and signals

Ville Pulkki and Matti Karjalainen

Department of Signal Processing and Acoustics
Aalto University, Finland

September 13, 2017

Sound as signal

In signal representations a physical or abstract variable is typically represented as a function of time, such as:

- Signal as a mathematical function:
  - Pure tone:
    $y(t) = A\sin(2\pi ft) = A\sin(\omega t)$
  - Random signal:
    $n(t) = \mathrm{rand}(t)$
- Discrete-time numeric sequence
  $x(n) = [0.1\ 2.2\ 3.5\ 4.0\ 3.1\ -0.9\ 2.1\ 0.5\ -1.1\ -2.1\ -0.8\ 0.2]$

Graphical presentations of signals

Sine wave

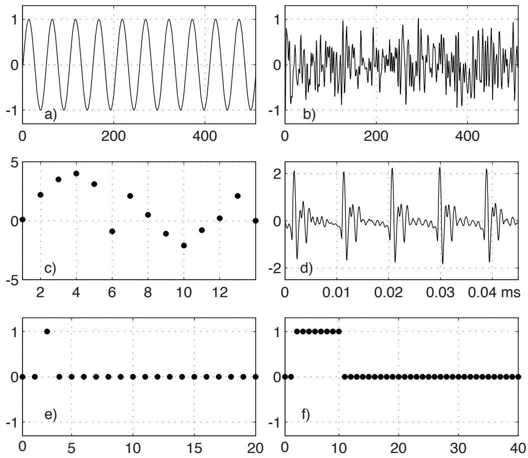
Random noise

Sample sequence

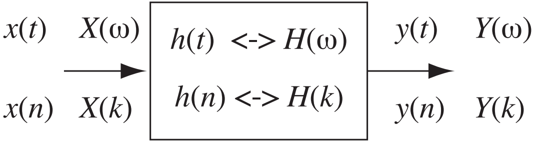
Speech waveform

Unit impulse

Unit pulse



Linear and time-invariant (LTI) systems



Properties of LTI systems

- Any (stable) LTI system can be fully represented by its impulse response
- Output cannot include any frequencies that are not in the input (no nonlinear distortion)
- Any bandlimited LTI system can be approximated by digital filters with arbitrary accuracy (theoretically)

Signal processing algorithms

- Convolution

$$y(t) = x(t) * h(t) = \int_{-\infty}^{+\infty} x(\tau)\, h(t-\tau)\, \mathrm{d}\tau$$

$$y(n) = x(n) * h(n) = \sum_{i=-\infty}^{+\infty} x(i)\, h(n-i)$$

- Fourier analysis

$$X(\omega) = \mathcal{F}\{x(t)\} = \int_{-\infty}^{+\infty} x(t)\, e^{-\mathrm{j}\omega t}\, \mathrm{d}t$$

$$X(k) = \mathcal{F}_{\mathrm{d}}\{x(n)\} = \sum_{n=0}^{N-1} x(n)\, e^{-\mathrm{j}k(2\pi/N)n}$$

Signal processing algorithms

- Fourier synthesis

$$x(t) = \mathcal{F}^{-1}\{X(\omega)\} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} X(\omega)\, e^{\mathrm{j}\omega t}\,\mathrm{d}\omega$$

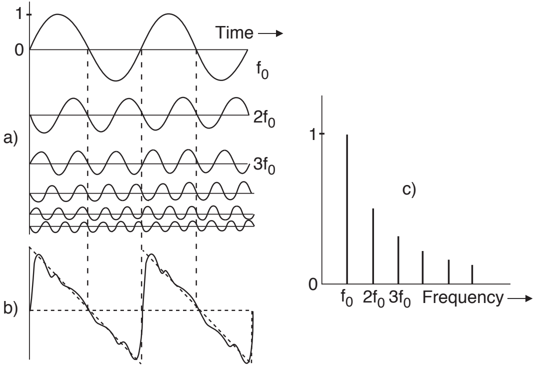
$$x(n) = \mathcal{F}_{\mathrm{d}}^{-1}\{X(k)\} = \frac{1}{N}\sum_{k=0}^{N-1} X(k)\, e^{\mathrm{j}k(2\pi/N)n}$$

- Convolution and Fourier transform

$$\mathcal{F}\{x(t) * y(t)\} = X(\omega)\cdot Y(\omega)$$

$$\mathcal{F}_{\mathrm{d}}\{x(n) * y(n)\} = X(k)\cdot Y(k)$$

Decomposition of sawtooth waveform



Spectrum analysis

- Magnitude spectrum

$$|X(\omega)|_{\mathrm{dB}} = 20\log_{10}|X(\omega)|$$

$$|X(k)|_{\mathrm{dB}} = 20\log_{10}|X(k)|$$

- Phase spectrum

$$\varphi(\omega) = \angle X(\omega) = \arg\{X(\omega)\}$$

$$\varphi(k) = \angle X(k) = \arg\{X(k)\}$$

- Phase delay $\tau_{\mathrm{p}}(\omega) = -\varphi(\omega)/\omega$
- Group delay $\tau_{\mathrm{g}}(\omega) = -\mathbf{d}\varphi(\omega)/\mathbf{d}\omega$

Fourier analysis with windowing

$$X(\omega) = \int_{t_\mathrm{b}}^{t_\mathrm{e}} w(t)\, x(t)\, e^{-\mathrm{j}\omega t}\, \mathrm{d}t$$

$$X(k) = \sum_{n=n_\mathrm{b}}^{n_e} w(n)\, x(n)\, e^{-\mathrm{j}k(2\pi/N)n}$$

- Rectangular window
- Hamming window
- Hann(ing) window
- Kaiser window
- Blackman (Blackman-Harris) window

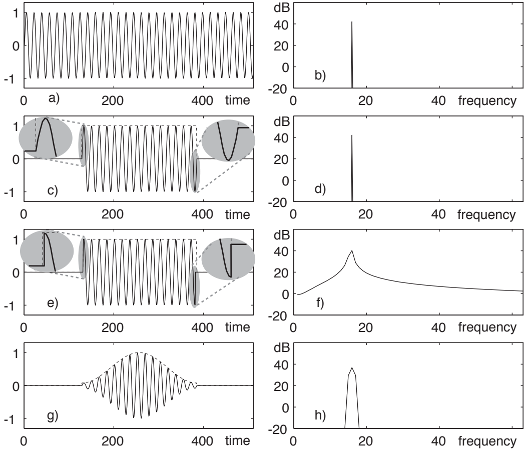
Spectrum analysis using Fourier analysis with windowing

Sine wave

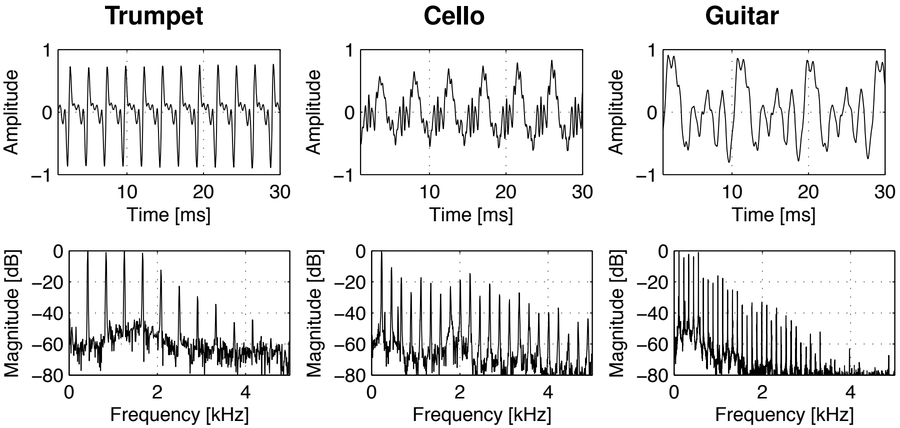
Sine wave windowed synchronously

Sine wave windowed non-synchronously

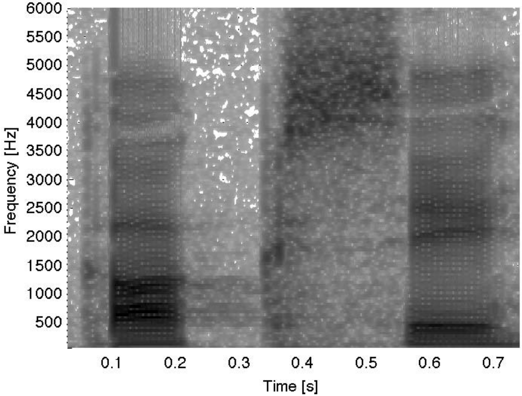
Sine wave Hamming-windowed



Magnitude spectra of example signals

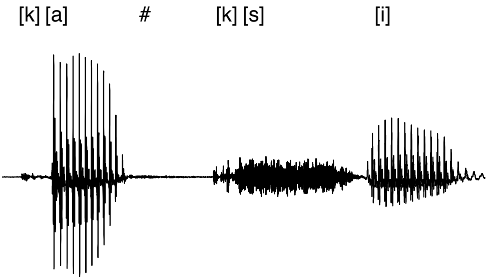


Spectrogram



Windowed Fourier magnitude spectrum shown for each time position

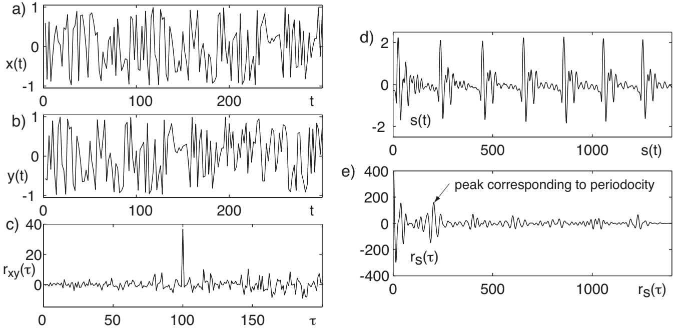
Corresponding waveform



Auto- and cross-correlation

$$r_{xy}(\tau) = \int_{-\infty}^{+\infty} x(t)\, y(t+\tau)\, \mathrm{d}\tau$$

$$r_{xy}(k) = \sum_{i=0}^{N-1} x(i)\, y(i+k)$$



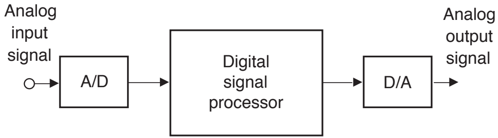
Cepstrum

$$c_x(t) = \mathcal{F}^{-1}\{\log |\mathcal{F}\{x(t)\}|\}$$

Commonly used in speech recognition as feature vector.

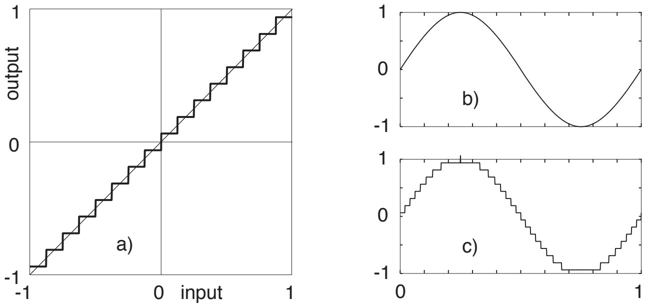
- Compute Fourier transform
- Logarithm of magnitude spectrum
- Inverse Fourier transform
- "Spectrum of the curve of magnitude spectrum"

Digital audio signal processing



- Analog-to-digital (A/D) converter
- Digital signal processor
- Digital signal processing (DSP) software
- Digital-to-analog (D/A) converter

Signal quantization in A/D conversion



- Linear quantization (PCM-coding)
- Discrete levels: 2^n (n= number of bits)
- 16–24 bits/sample in audio ($\geq$ 96 dB SNR)
- Sample rate: e.g. 44100 or 48000 samples/s

Filtering

Filters are DSP components that have frequency-dependent magnitude and/or phase response. Needed often in audio techniques.

- low-pass filter, attenuate high frequencies above cutoff frequency
- high-pass filter, attenuate low frequencies
- band-pass filter, attenuate low and high, and leave a band unmodified
- band-reject filter, correspondingly
- all-pass filter, modify only phase response
- arbitrary-response filter, design the response as needed for each frequency

Z-transform

Linear transform of sequence

$$X(z) = \mathcal{Z}\{x(n)\} = \sum_{n=-\infty}^{\infty} x(n)\, z^{-n}$$

Unit delay as building element:

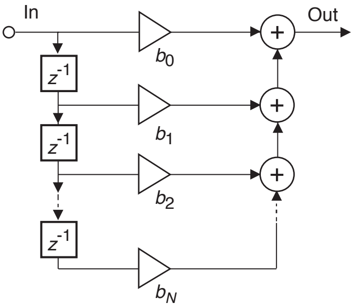
$$\mathcal{Z}\{x(n-1)\} = z^{-1} X(z)$$

Digital filtering can be expressed as a polynomial of z^{-1}

Digital filtering: FIR filters

FIR = Finite impulse response

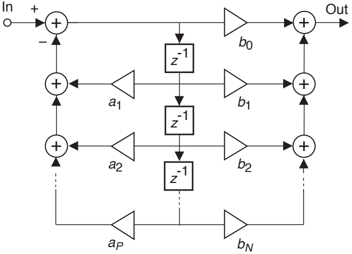
$$H_{\mathrm{FIR}}(z) = \sum_{n=0}^{N-1} b_n\, z^{-n} = b_0 + b_1\, z^{-1} + \cdots + b_{N-1}\, z^{-(N-1)}$$



Digital filtering: IIR filters

IIR = Infinite impulse response

$$H_{\mathrm{IIR}}(z) = \frac{\sum_{n=0}^{N-1} b_n\, z^{-n}}{1 + \sum_{p=1}^{P-1} a_p\, z^{-p}} = \frac{b_0 + b_1\, z^{-1} + \cdots + b_{N-1}\, z^{-(N-1)}}{1 + a_1\, z^{-1} + \cdots + a_{P-1}\, z^{-(P-1)}}$$



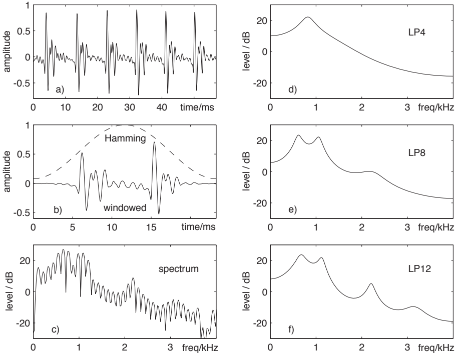
Linear prediction (AR-modeling)

Modeling of signal generation with flat spectrum excitation (impulse or noise) and IIR (all-pole) filter.
Speech example:

Signal

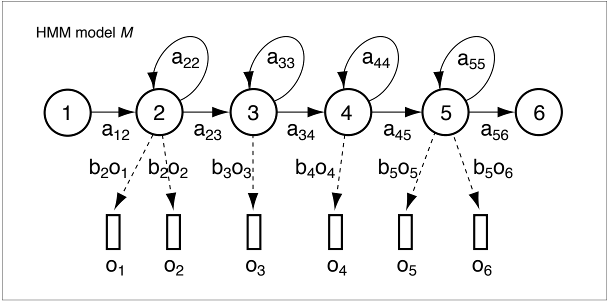
Windowed

FFT-spectrum



LP-spectra

Hidden Markov models (HMM)



- For probabilistic modeling of state sequences
- Used especially in speech recognition and synthesis

References

These slides follow corresponding chapter in: Pulkki, V. and Karjalainen, M. Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics. John Wiley & Sons, 2015, *where also a more complete list of references can be found*.