Aalto University
School of Electrical
Engineering

Communication acoustics
Ch 15: Time–frequency-domain processing and coding of audio

Ville Pulkki, Juha Vilkamo

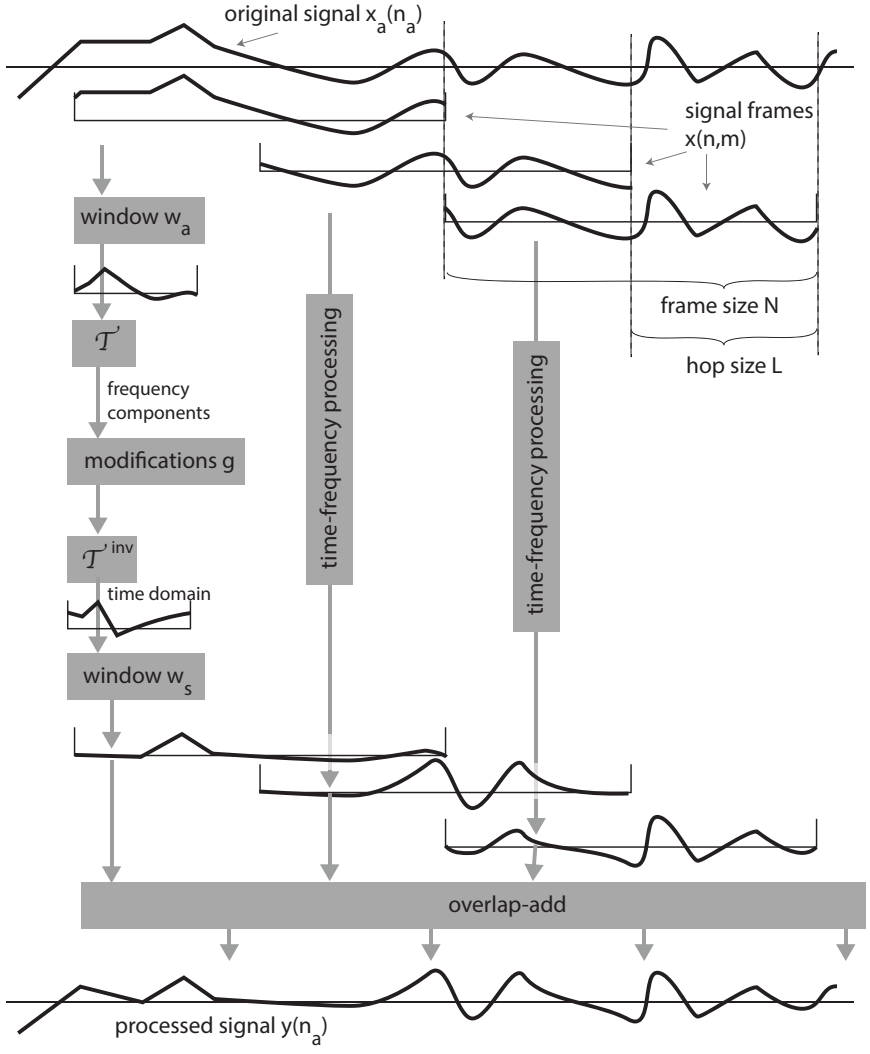
Department of Signal Processing and Acoustics
Aalto University, Finland

December 12, 2017

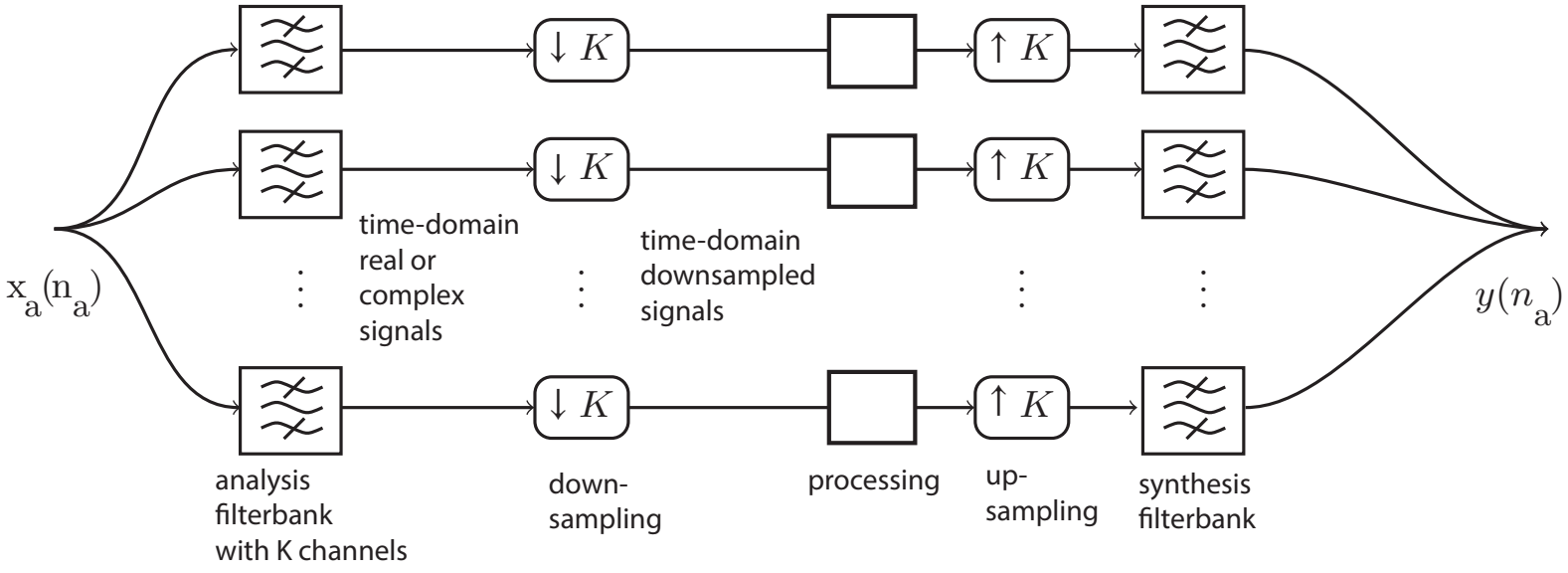
Time-frequency processing of audio

- Human hearing is more or less based on separate processing of different frequency bands
- Trend in audio is to process audio signals also in frequency bands
- Perceptual coding of audio
- Spatial audio
- Hearing aids

Frame-based processing

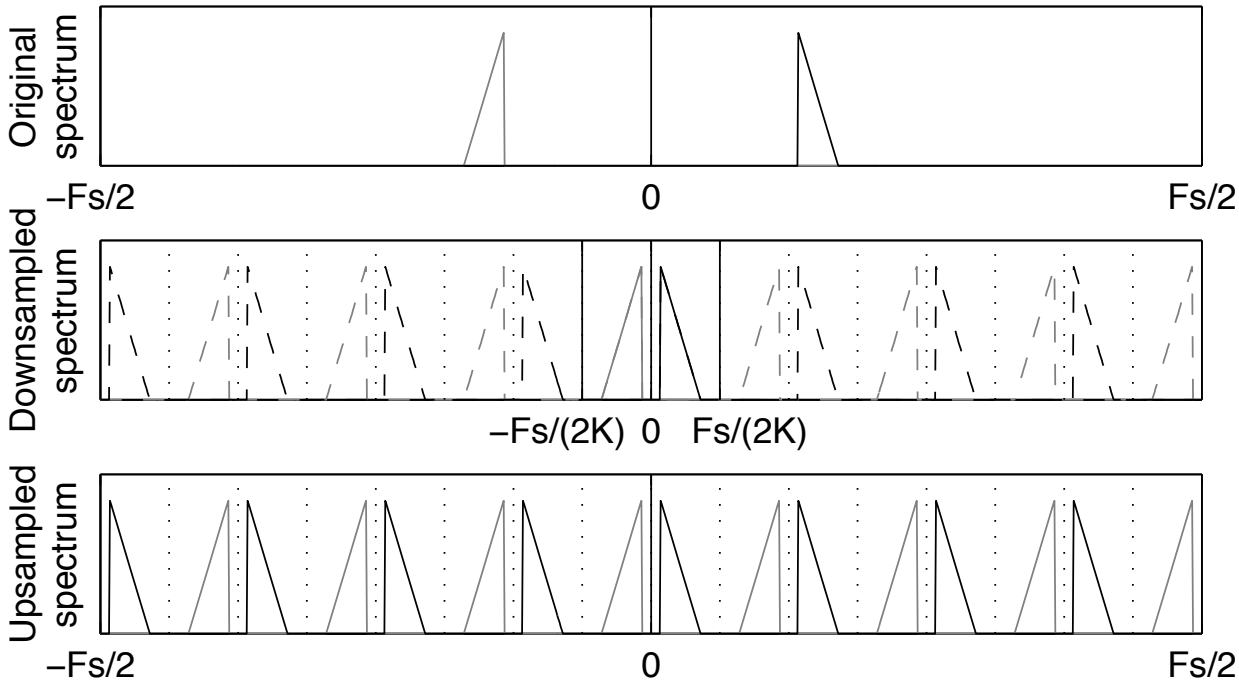


Downsampled filterbank processing



- K filters produce K time-domain signals
- Downsampled by factor K (every K:th sample is preserved)
- Upsampled by factor K (add $K - 1$ samples after each sample)
- Total bit rate can be doubled to obtain robust features (audio processing) or not changed (audio coding)
- Actual systems employ efficient mathematically equivalent structures

Downsampling-upsampling in filter banks

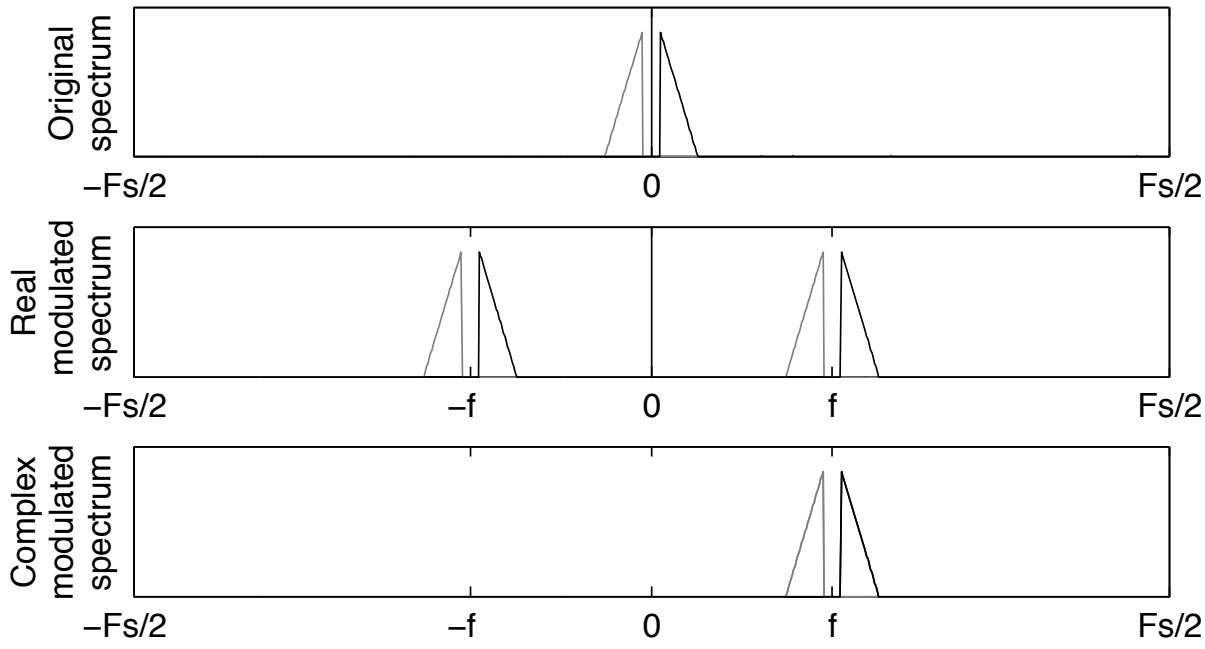


- Downsampling maps the band-pass filtered signal to the lower frequencies, and upsampling reconstructs the original spectrum
- Extra components are suppressed by the synthesis filter

Modulation with tone sequences

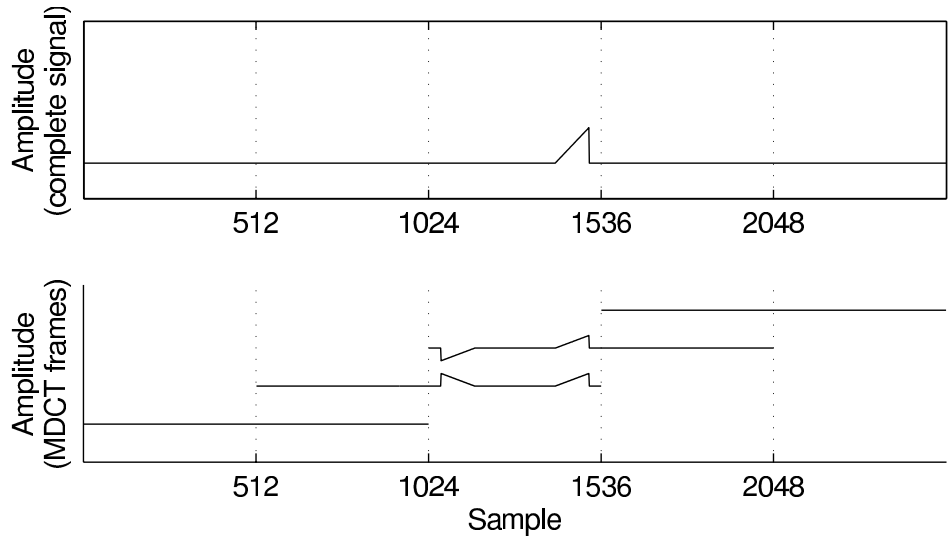
- Modulation is used as a tool to shift a spectrum of a signal or a filter.

$$x_{\mathrm{modR}}(n) = x(n)\cos(\omega n) = x(n)\frac{e^{j\omega n} + e^{-j\omega n}}{2}$$



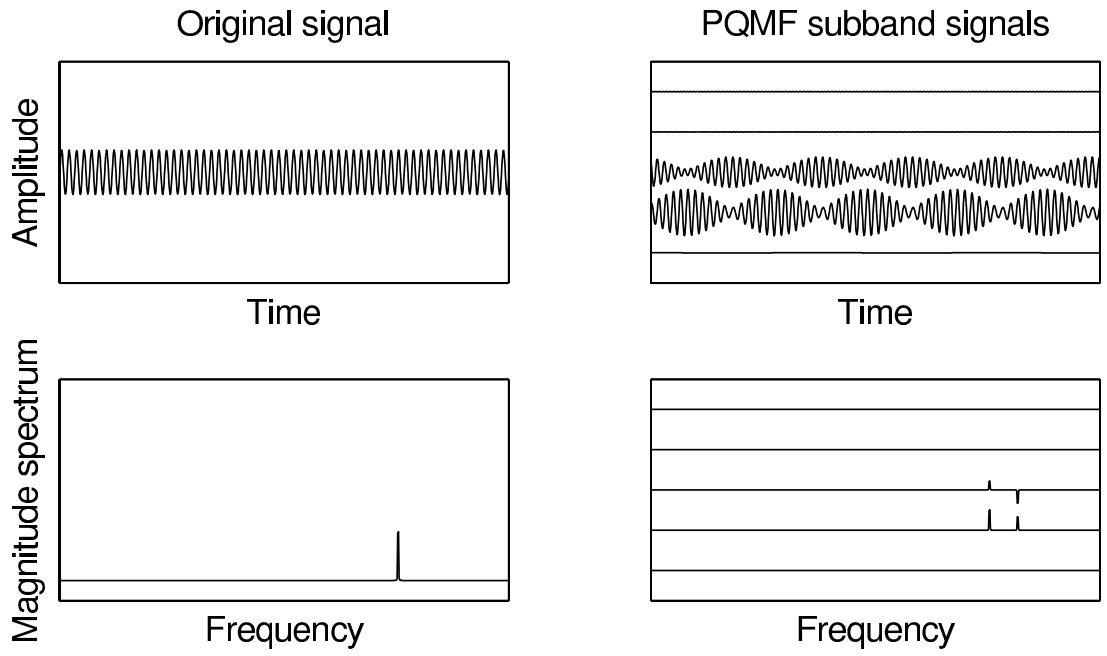
Aliasing

- A non-linear effect where signal components appear where none existed in the original signals
- Frequency aliasing, time aliasing
- In MDCT, time-domain aliasing may appear, if frames are processed differently



Aliasing

- In PQMF, frequency-domain aliasing may appear, if frequency bands are processed differently



Time-frequency transforms

Definitions:

- *Critical sampling*: combined sampling rate of the transformed frequency bands is the same as that of the original signal. A feature audio coding transforms
- *Oversampling*: combined sampling rate of the transformed bands is higher than that of the original signal. A feature of robust audio processing transforms
- *Perfect reconstruction* means that the original signal waveform is retrieved exactly after the inverse transform
- *Near-perfect reconstruction*: original signal can be retrieved with a high degree of accuracy, but not exactly.

Short-time Fourier transform

- Analysis window sequence $w_a(n)$ is applied to a signal sequence $x(n)$

$$X(k) = \sum_{n=0}^{N-1} w_{\mathrm{a}}(n)x(n)e^{-j2\pi kn/N}$$

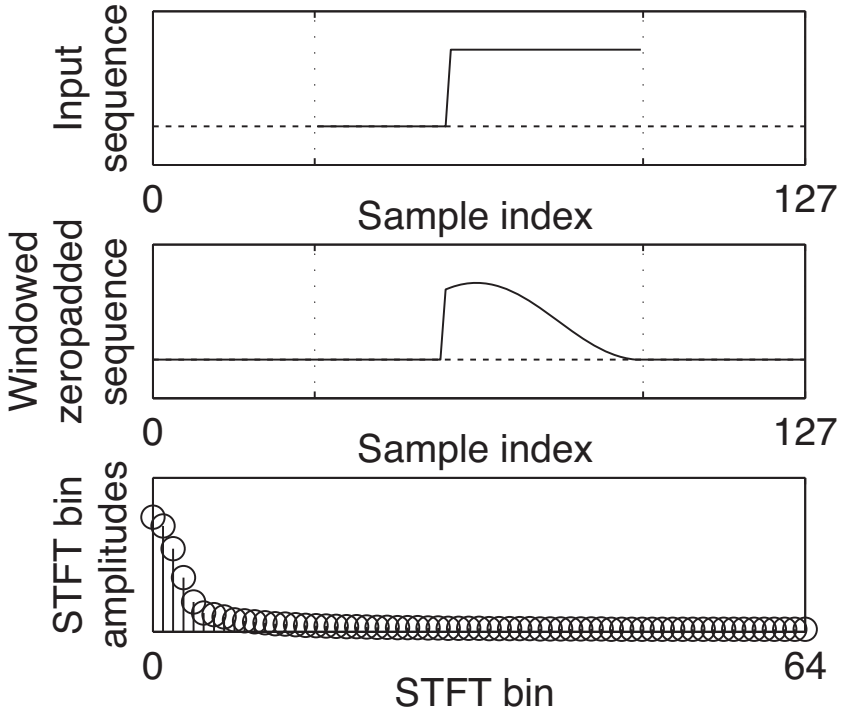
- In synthesis window sequence $w_s(n)$ is applied to $X(k)$

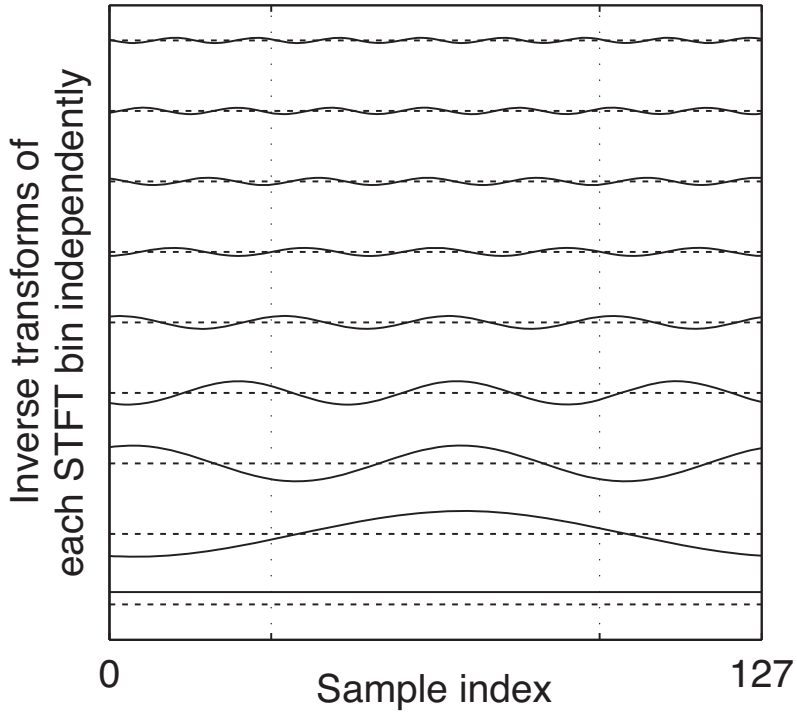
$$y(n) = \frac{1}{N} w_{\mathrm{s}}(n) \sum_{k=0}^{N-1} X(k)e^{j2\pi kn/N}$$

- Overlap-add processing

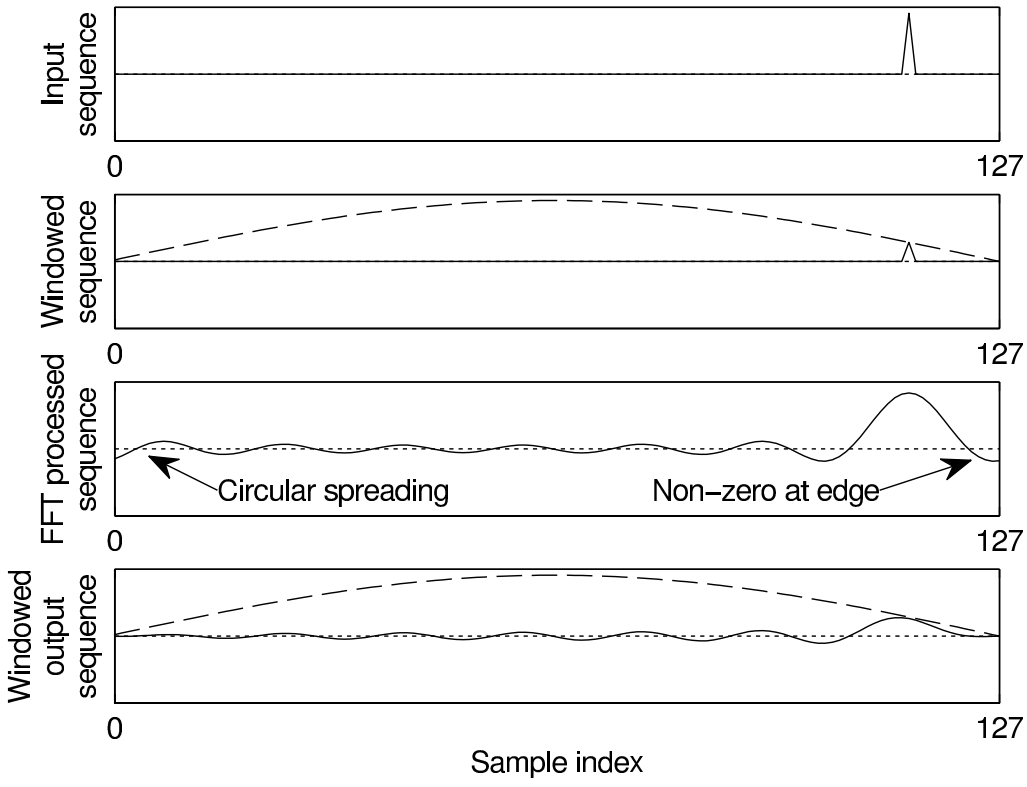
STFT

- Windowed signal is presented with circularly continuous tones





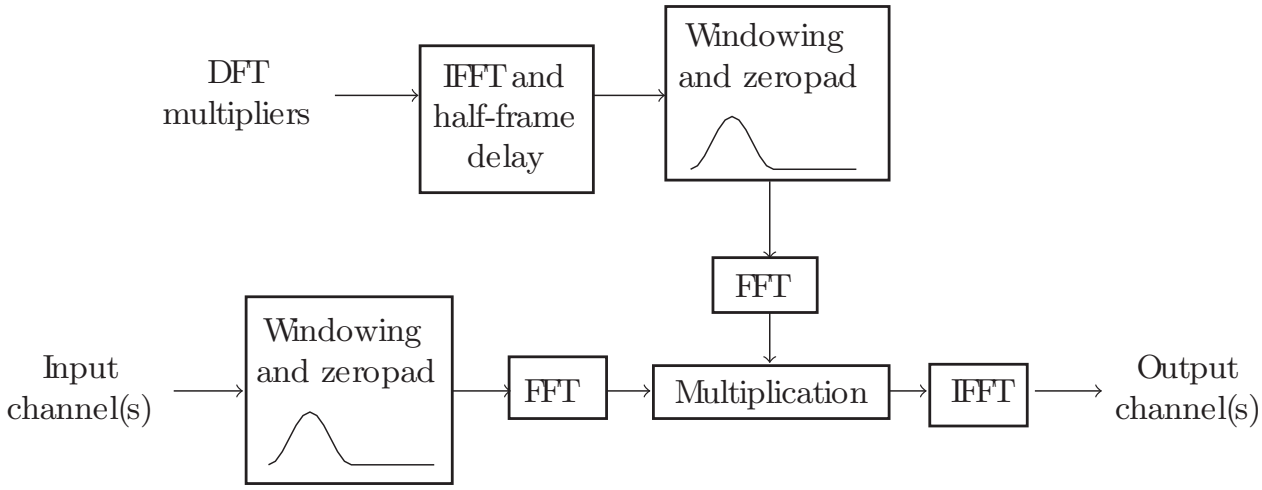
Effect of synthesis window in STFT



- Avoid at least partly temporal aliasing
- If signal is processed heavily in frequency domain, aliasing artifacts may still occur

Alias-Free STFT

- A method for pre-processing the time-frequency processing coefficients such that aliasing does not occur
- Processing coefficients are transferred to time domain, and windowed there before applying to signal
- Artifacts avoided, higher CPU load than in STFT



Modified Discrete Cosine Transform (MDCT)

- Perfect reconstruction, real-valued, and critically sampled transform
- Widely used in audio coding due to non-redundancy and its property of representing narrowband signals with a relatively small number of prominent spectral coefficients

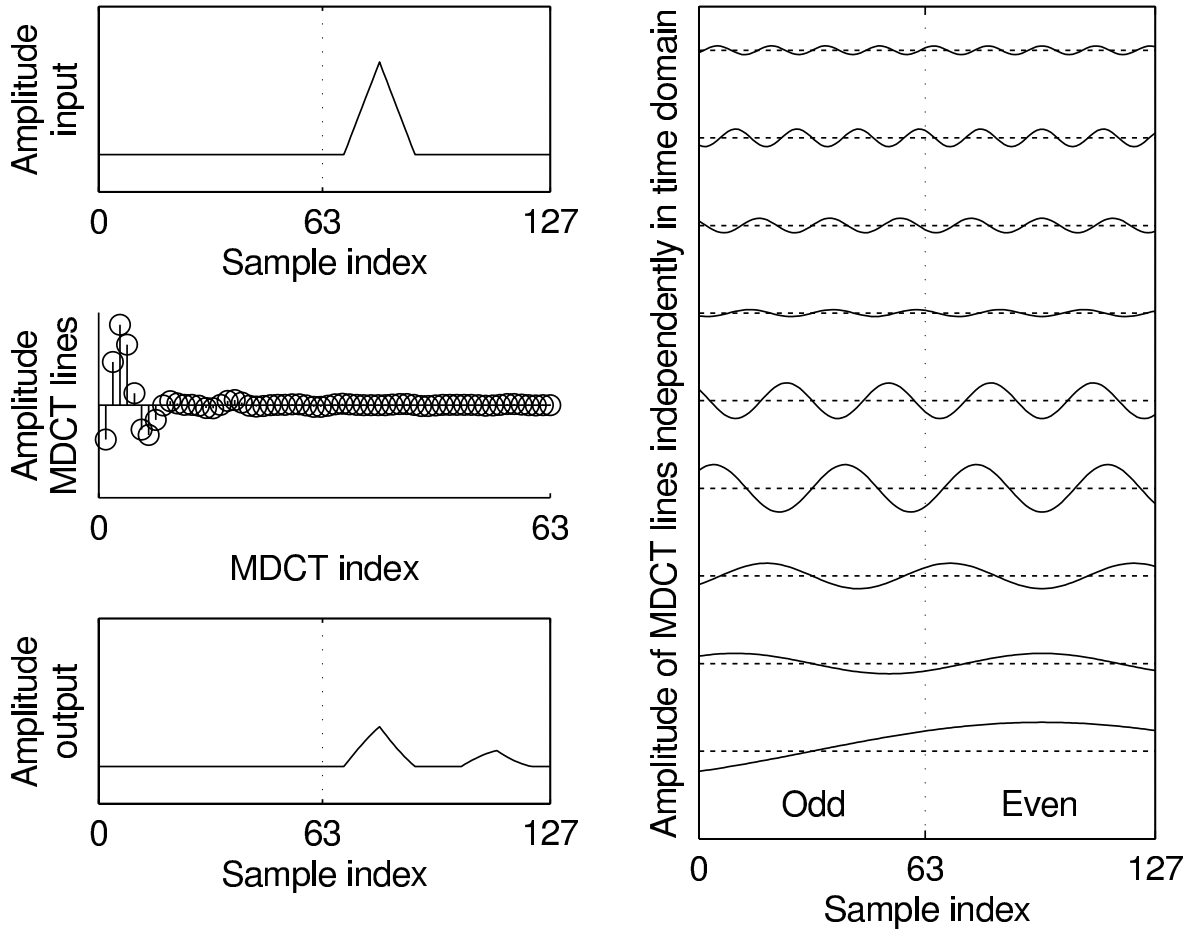
$$X(k) = \sum_{n=0}^{2N-1} w_a(n) x(n) \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2} + \frac{N}{2}\right)\left(k + \frac{1}{2}\right)\right]$$

$$y(n) = \frac{2}{N} w_s(n) \sum_{k=0}^{N-1} X(k) \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2} + \frac{N}{2}\right)\left(k + \frac{1}{2}\right)\right]$$

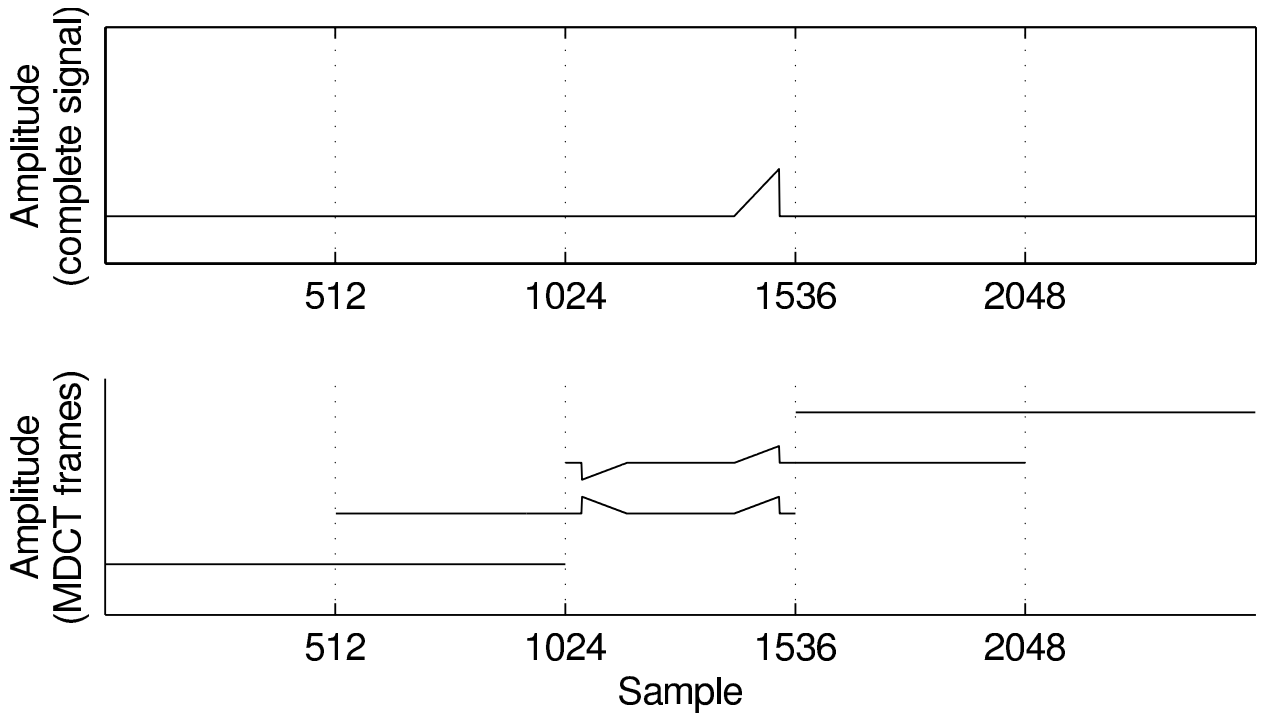
Modified Discrete Cosine Transform (MDCT)

- Half-overlapping frames (windows of N samples the hop size is $N/2$ samples)
- Frequency partials of the MDCT represent sinusoids that are odd symmetric in the first half and even symmetric in the second half of the frame

Modified Discrete Cosine Transform (MDCT)



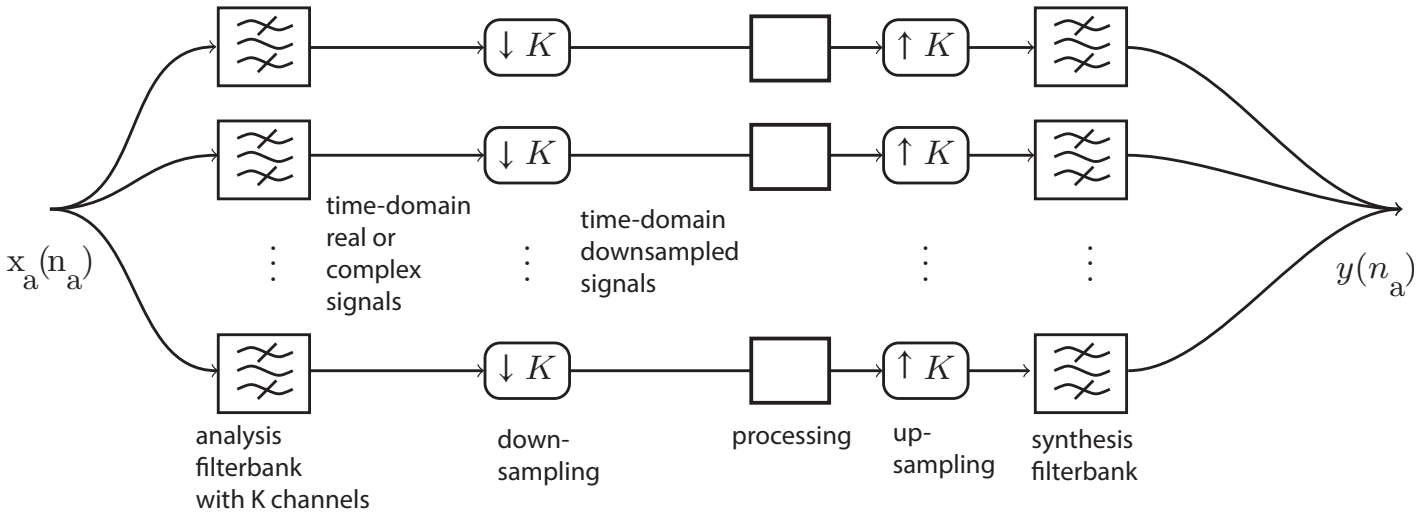
MDCT



- Overlap-added sub-sequent frames cancels time-domain aliasing.
- Aliasing-cancellation works in audio coding (signal not changed prominently), while fails for audio processing

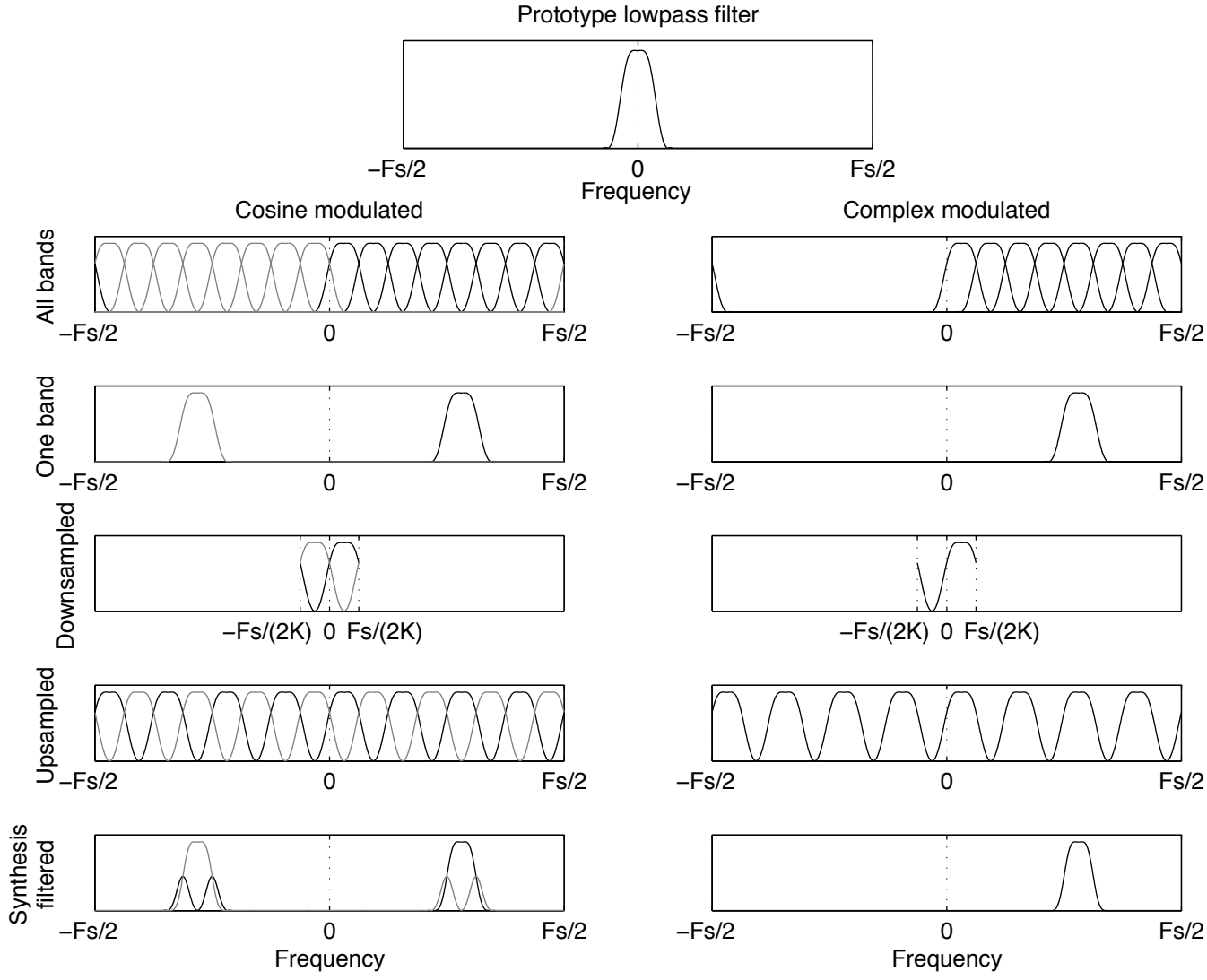
Pseudo-Quadrature mirror filter (PQMF) bank

- Real-valued, critically sampled filter bank
- Synthesis and analysis filters are implemented as FIR filters which are modulated to different frequencies
- Near-perfect reconstruction



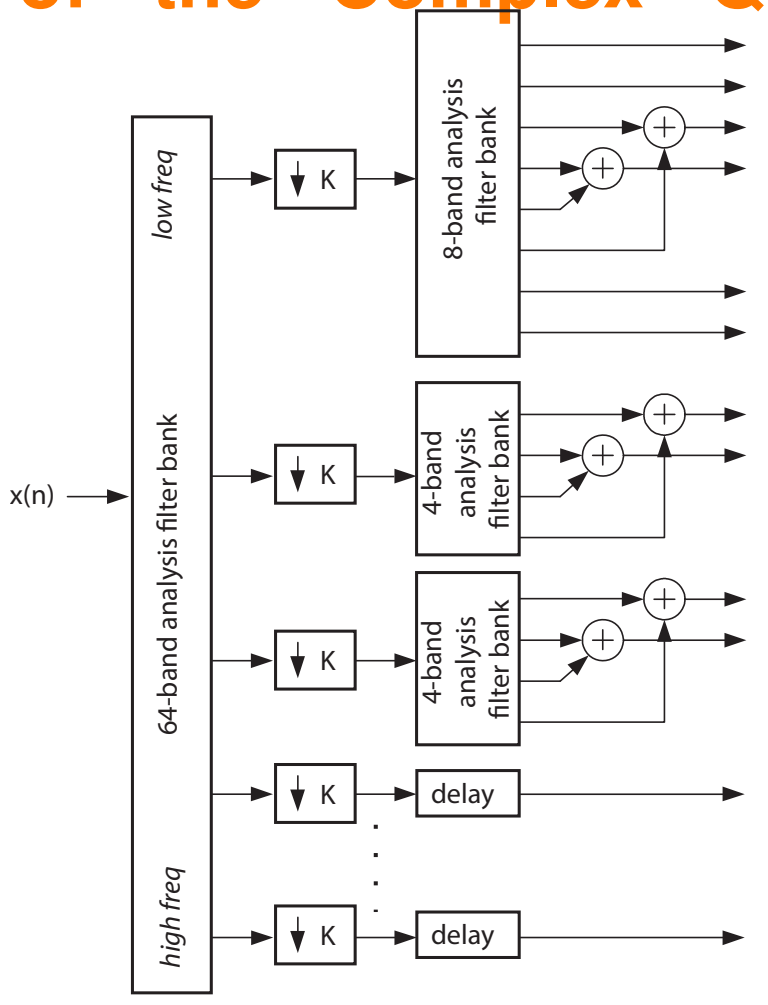
Complex QMF

- PQMF is prone to artifacts, it is an audio coding filter bank
- Complex QMF, use complex modulators in analysis and synthesis filter design
- Double-oversampled, redundant representation
- Aliasing becomes negligible for any kind of processing
- Used in time-frequency-domain spatial audio processing

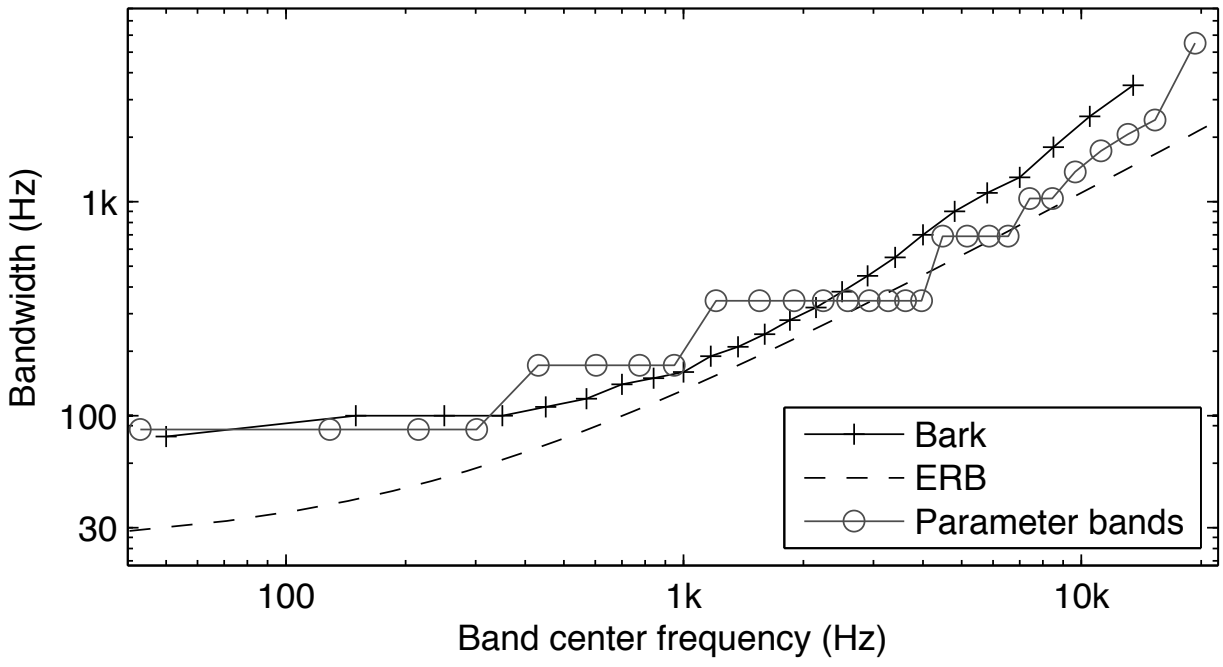

Prototype lowpass filter
−Fs/2
0
Fs/2
Frequency
Cosine modulated
Complex modulated
All bands
One band
Downsampled
Upsampled
Synthesis filtered
−Fs/2
0
Fs/2
−Fs/(2K) 0 Fs/(2K)
Frequency

Sub-sub-band filtering of the Complex QMF bands

- Complex QMF is robust for aliasing
- Frequency resolution is linear
- Rough estimation of critical bands by dividing QMF bands to sub-bands



Hybrid QMF sub-sub-band filter bank

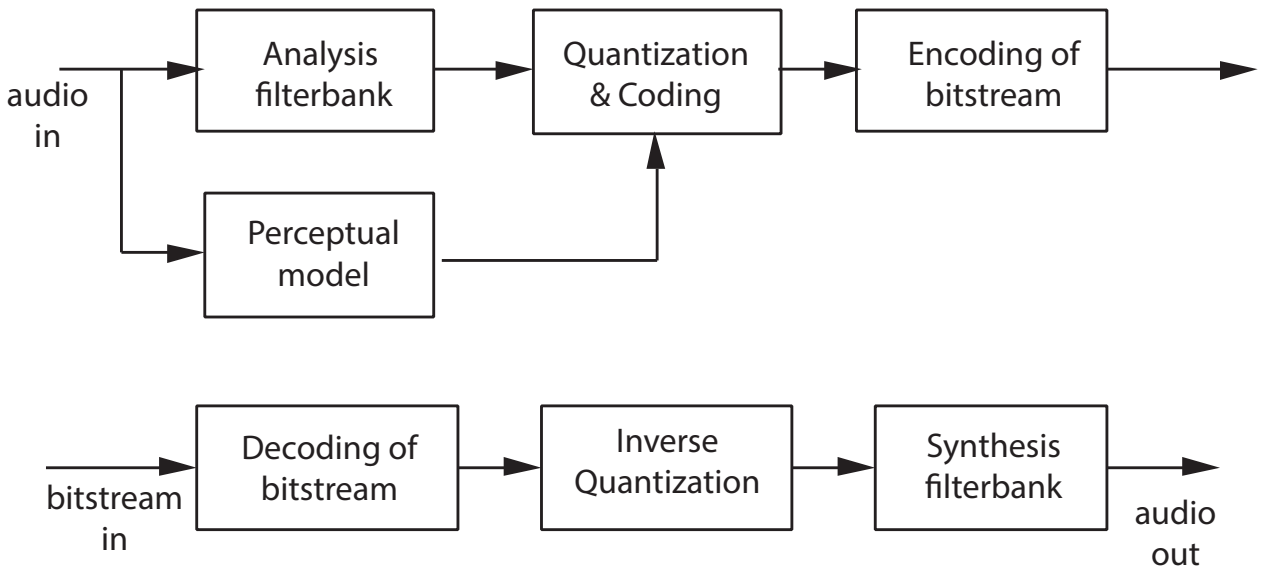


Time–frequency-domain audio-processing techniques

- Audio coding
  - Masking-based
  - Spectral band replication
  - Parametric Stereo, MPEG Surround, and Spatial Audio Object Coding
- Stereo upmixing and enhancement for loudspeakers and headphones
- Microphone array processing techniques
  - Spatial sound reproduction
  - Beamforming

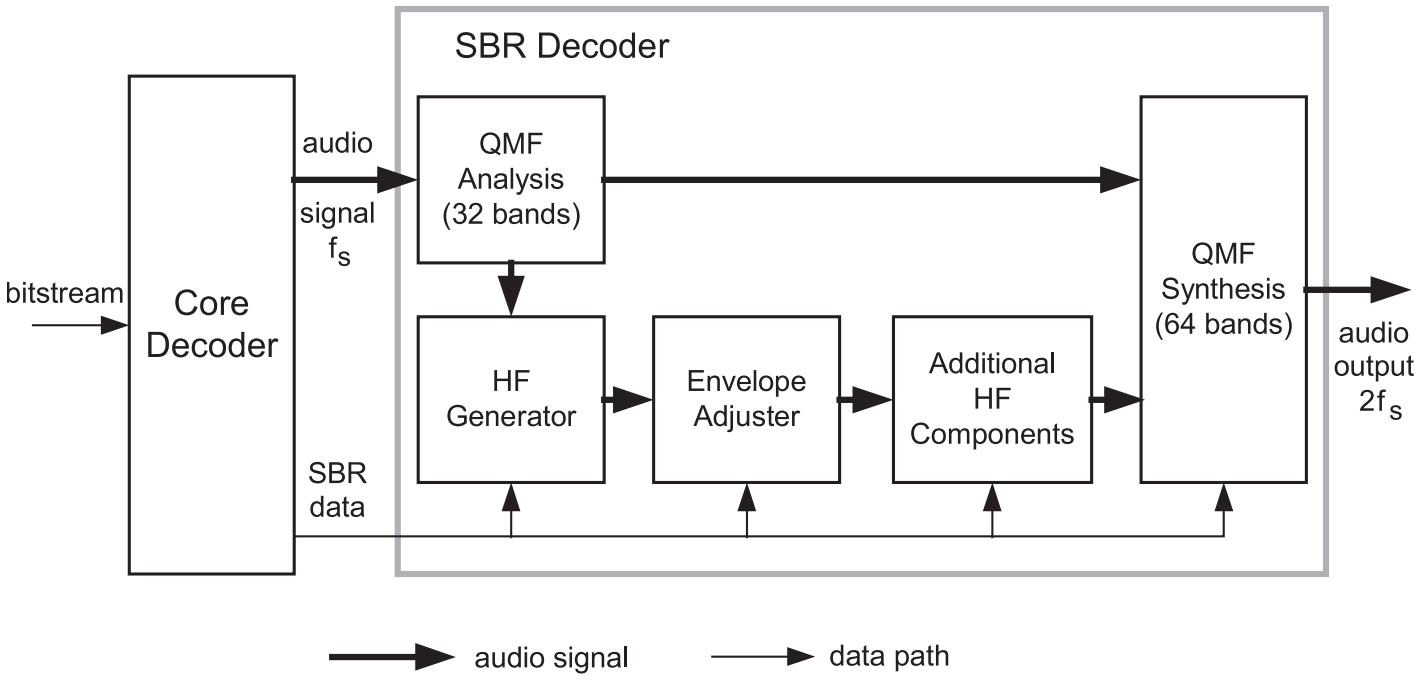
Masking-based audio coding

- Estimate masking threshold with auditory model
- Shape the spectrum of quantization noise accordingly
- MPEG-1 layer 3 (MP3), MPEG-2, AAC



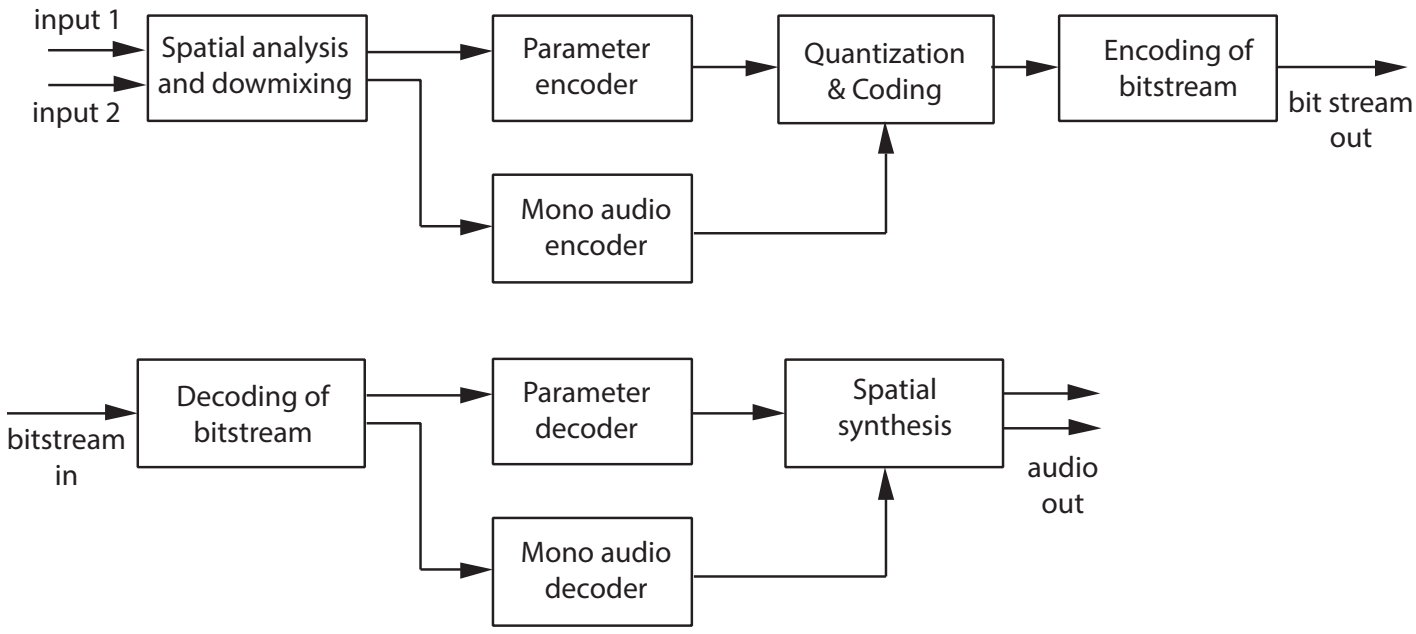
Spectral band replication

- Send only bands at low frequencies
- Predict high frequencies from transmitted low frequencies



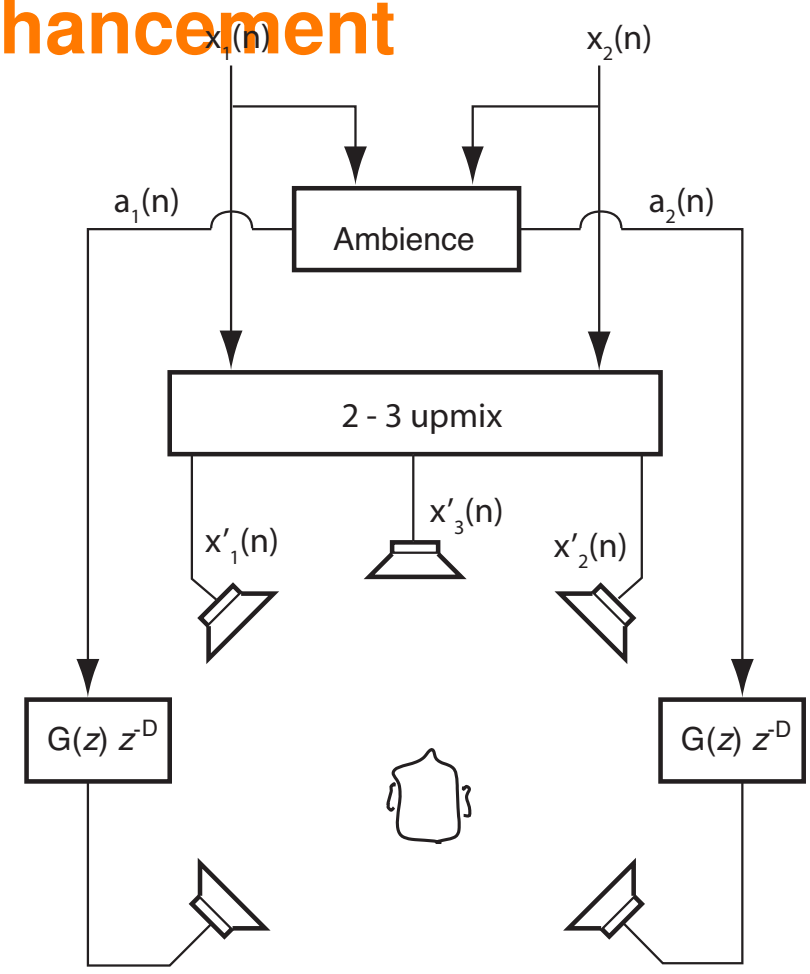
Multichannel audio coding

- Parametric Stereo, MPEG Surround, and Spatial Audio Object Coding
- Estimate the differences btw input channels in time-frequency domain
- Send downmix audio + metadata, and reproduce sound using them



Stereo upmixing and enhancement

- How to produce 5.1 audio from stereophonic content?
- Analyze properties of stereo signals in TF-domain
- Surround reproduction of "ambient" or "diffuse" sounds, while keeping "direct" sounds in the front



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.*