Collins: Introduction to Computer Music

- Dekker grunnleggende elementer av musikalsk lyd, både som signal og som opplevde sansekvaliteter

- Dekker grunnleggende prinsipller for opptak, lagring, representasjon, analyse og bearbeidelse av musikalsk lyd

- Derfor: Både grunnforståelse og verktøy til bruk i videre lydanalysearbeid

- Strategi: Veksle mellom teoretisk forståelse og praktisk erfaring med programvare

Collins: Introduction to Computer Music

1. Introduction
2. Recording
3. Analysis
4. Processing
5. Synthesis

Chapter 1, Introduction

1.1 What is Computer Music?

- Computer music = "music that involves a computer at any stage of its life cycle"
- With digital devices everywhere, computer music everywhere
- Sound source may be anything from outside the computer (instruments, environment) to inside the computer (various synthesis models) and combinations (sound processing)
- Computer music eminently interdisciplinary, combining art and science (and also natural and human sciences): music, psychology, biology, math, acoustics, physics, informatics, etc.

1.1.1 Some Examples of Computer Music

- Comment: *electroacoustic music* (EA music) traditionally much wider than computer music, e.g. the *musique concrète* and *synthetic* EA music of the 1950s
- Groundbreaking work in digital music synthesis by Risset and Chowning in the 1960s (and Knut Wiggen in the 1970s)
- Max Matthews and the *Music x* software development
- Explosion of digital music production in the 1980s and following decades

1.1.2 Sociable Computer Musicians

- Some interesting reflections on social aspects of computer music:
- Interactive music creation
- Music sharing
- Radical changes to both production and distribution
- Horizions: maybe blurred boundaries between composer, musician and listener by various notions of *active music* and *adaptable music*

1.1.3 Why Investigate Computer Music?

- Advantages of digital sound technology: capturing, storing, processing, representing, diffusion, etc.
- Modeling musical sound
- Radical new forms of musical expression
- Useful tools for composition
- Possibilities of making ergonomically 'impossible' music (e.g. extremely fast drumming)
- However: many open questions on the relationships between sound technology and music perception as well as aesthetics
- Schaeffer: the *acousmatic* perspective, meaning considering sound regardless of origin

1.2 Quickstart Guide to Computer Music.

1.2.1 Sound Waves and the Brain.

- Sound waves
- Transduction air-ears-brain-mind, have a decent understanding of the different stages here, hence
- Physics: understanding the generation and propagation of sound
- Psychoacoustics: how sound is transformed in our percption
- Neurobiology: the biological bases for sound perception
- Cognitive psychology: how we make sense of sound
- But also other branches of psychology, e.g. emotions, pleasure/displeasure, etc.

1.2.2 The Time Domain

- In general: graphical plotting gives important insights
- Plotting time-varying signal
- Waveform
- Time-domain
- Range: normalize to -1.0 to 1.0
- Amplitude
- Power
- Root mean square (RMS) $x_{\mathrm{rms}} = \sqrt{\frac{{x_1}^2 + {x_2}^2 + \cdots + {x_n}^2}{n}}.$
- Operations on a signal:

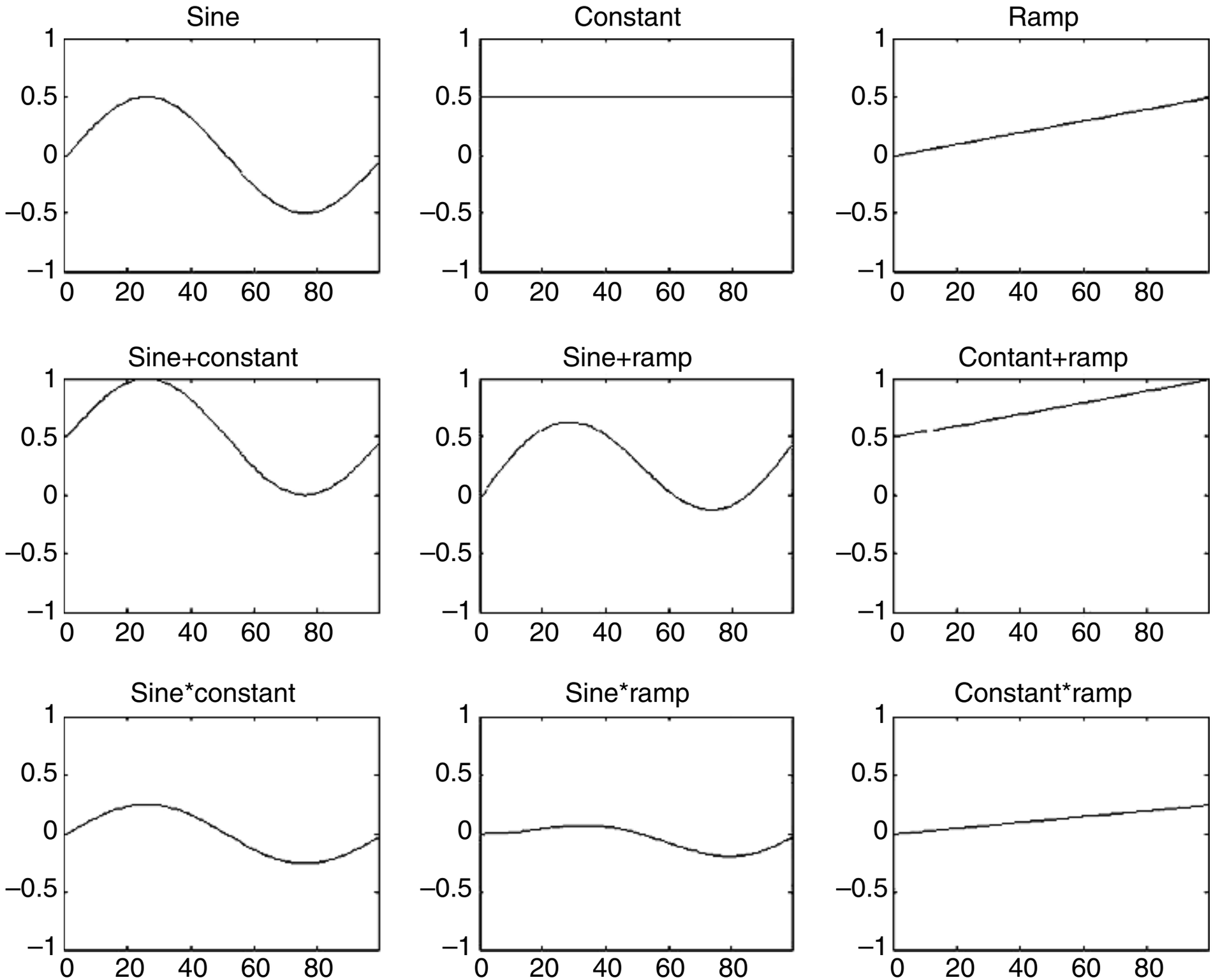


Figure 1.1 Basic signal operations in the time domain. The three signals on the top row are combined by addition in the middle row and multiplication in the lowest row. The signal operations act pointwise over time. The 'constant' signal is also called a DC (direct current) offset. When it is added to another signal, it offsets that signal; when it multiplies, it scales the range.

1.2.2 The Time Domain

- Linear
- Logarithmic
- General point: mapping a feature dimension to some scale and various principles of scaling
- Try out for yourself with e.g. the *Grapher* (on the Macintosh) or similar software. Very useful to visualize these basic operations!
- Decibels
- Loudness

Human hearing is approximately logarithmic (see sidebar) over much of its scope. For this reason, it is convenient to convert amplitude to units known as **decibels**.[4] The standard equation for this is:

$$\text{value in decibels} = 10 \log_{10} \left(\frac{\text{input power}}{\text{reference power}} \right) \tag{1.1}$$

where the power of a signal is the square of the amplitude, and $\log_{10}$ is used for logarithms to base 10. To work with amplitude directly:

$$\text{value in decibels} = 20 \log_{10} \left(\frac{\text{input level}}{\text{reference level}} \right) \tag{1.2}$$

The logarithm must be taken to a base, 10 in this case, and the measurement of decibels is always with respect to a reference level. If we imagine signals within the range −1 to 1, we could set a reference level at full power of 1, or make it very small, perhaps 10^{-12} watts per square meter (this latter case is often denoted by dB SPL, which stands for sound pressure level). In the first case, full amplitude would be 0 dB and all lesser amplitudes would be represented by negative numbers of decibels. In the second, 1 would be 120 dB ($10 \log_{10} 1/10^{-12}$) relative to the reference. In both cases, an amplitude of zero would be negative infinity, since that is the logarithm of zero. You may have seen audio software and hardware meters using various conventions here; a mixer's working scale could be from +9 dB down to infinity, allowing some 'headroom' rather than immediate overloads. I have also skipped a couple of technicalities – for example, taking an average measure of amplitude during some time span – here for the sake of the exposition [Loy, 2007a, Chapter 4].

1.2.3 Periodicity

- Periodicity = that which is repeated
- Oscillation = repeated motion back and forth, up and down, etc.
- Motion around a circle, or rotation, the basis for understanding musical acoustics
- The so-called 'unit circle' with size of radius, speed of radius rotation, and position/angle of radius = the basis for studying sound
- Trigonometric relations of sine and cosine
- Complex numbers (more on this later)
- Basis for Fourier analysis

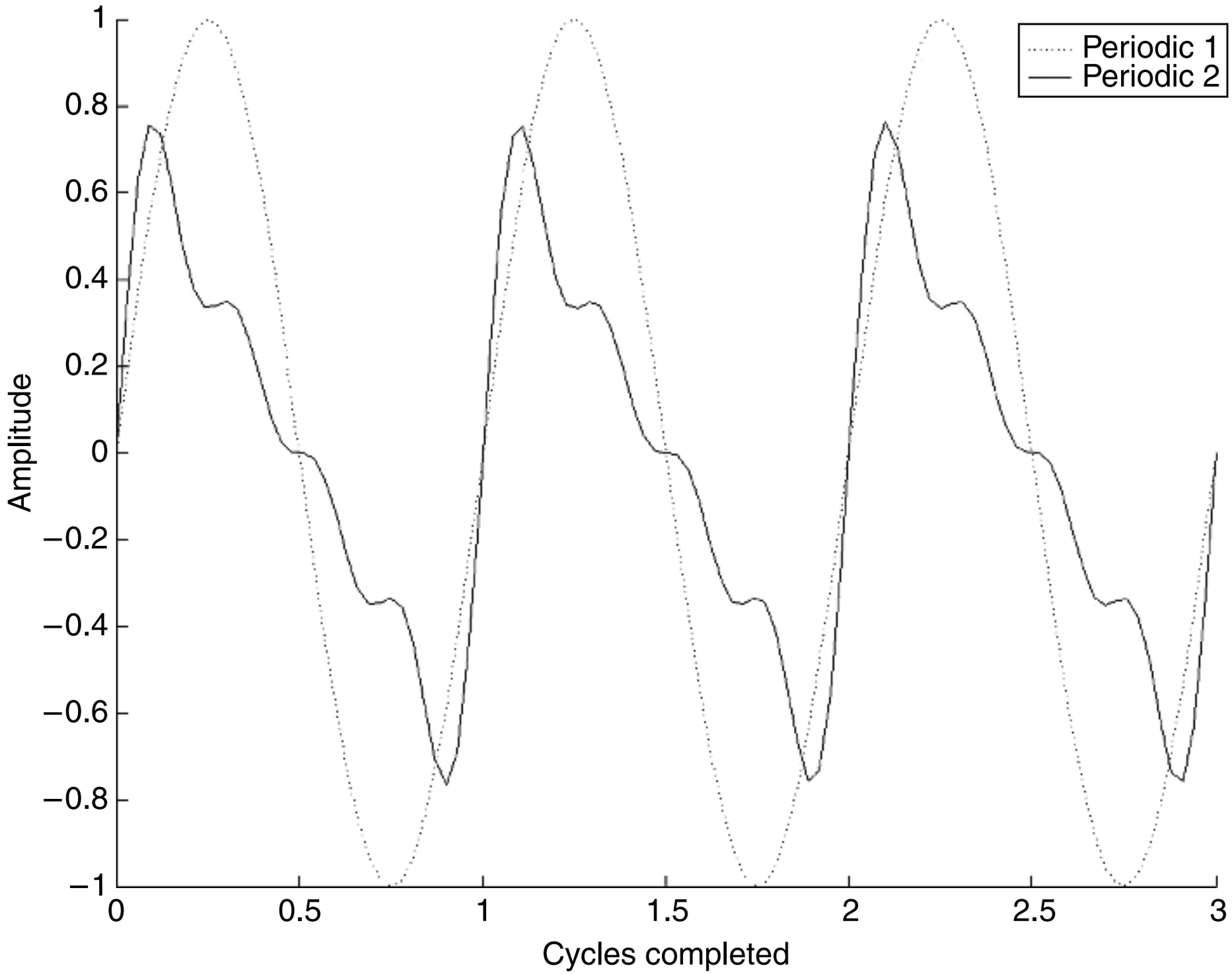


Figure 1.2 Periodic waveforms.

1.2.3 Periodicity

- Sine tones
- Fundamental frequency
- Partials, harmonics, overtones
- Pitch
- Harmonic sounds
- Inharmonic sounds
- Additive synthesis
- Spectral analysis

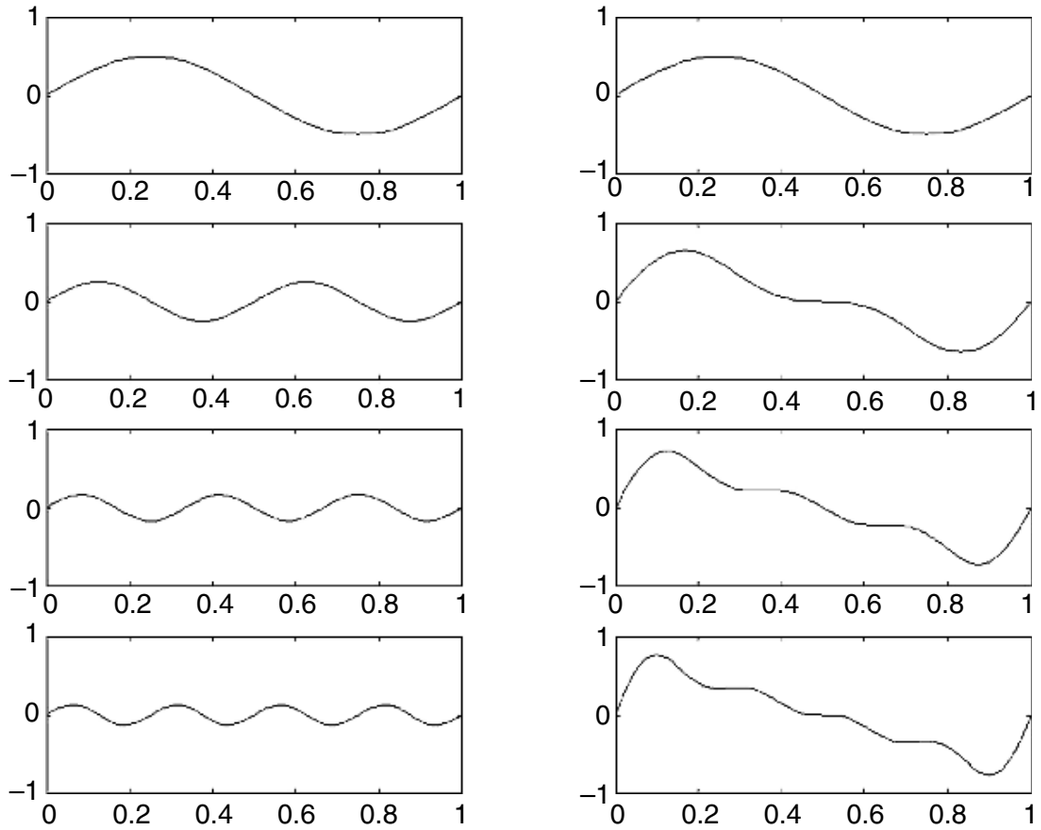


Figure 1.3 Sinusoidal components of a periodic sound. The left column shows each individual sine component (they vary in their amplitude). The right column gives the mix so far at each stage, as the sines are added together down the page.

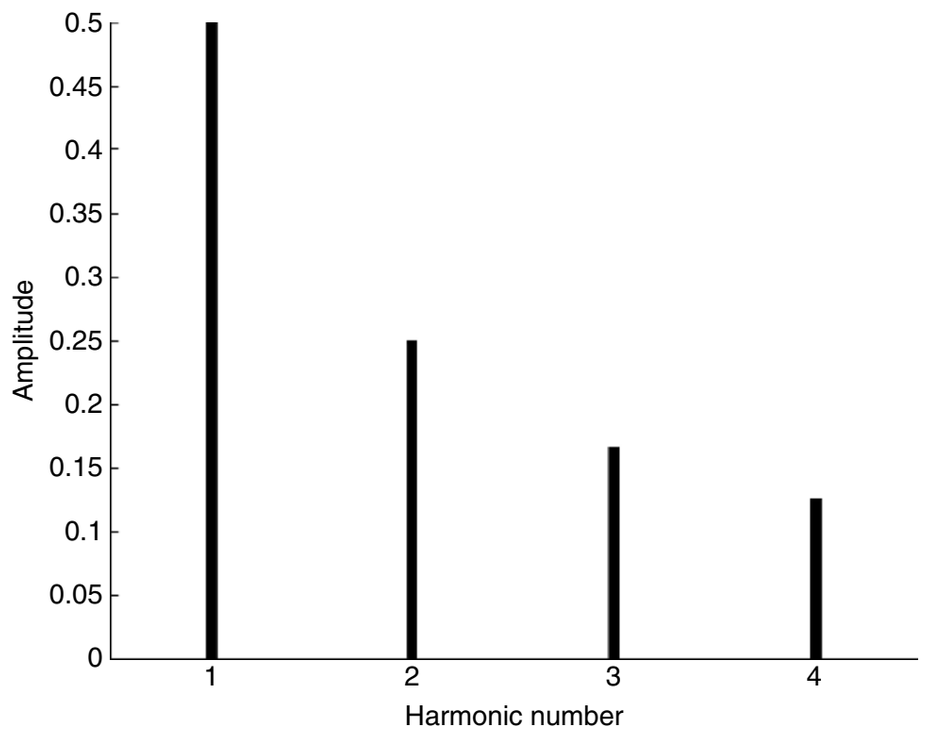


Figure 1.4 Line spectra. The frequency and amplitude of each component of the periodic sound are indicated (though phase information is dropped).

1.2.4 The Frequency Domain

- Most natural sounds pseudo-periodic (or quasi-periodic)
- Steady state segments in sounds
- Different modes of vibrations in addition to the fundamental = the basis for spectrum
- Spectral analysis = testing the energy level of different frequencies
- Fourier analysis: testing an infinitely long signal
- Short-time Fourier transform (STFT): testing a time-limited signal
- A 'window' is used to extract a signal segment

1.2.4 The Frequency Domain

- On discretized sound, window size is measured in number of samples
- Time-frequency tradeoff (more on this later)
- Effects of window shape on the spectrum
- Effects of 'hop size' on the spectrum
- Discrepancies between the seen and the heard
- Fast Fourier Transform (FFT) = an efficient algorithm needing power of 2 number of samples
- Spectrogram = a series of FFTs that show the evolution of frequency amplitude over time
- Phase information may be hidden, but needs to be recovered for IFFT and other operations (e.g. in a *digital vocoder*)

1.2.5 Digital Audio

- Discretization = the basis for digital audio
- All information in the form of bits
- General question of *resolution* in digital audio (and other fields)
- Bit resolution and temporal (sampling rate)
- Analog to digital conversion
- Number of bits determine signal-noise ratio
- Sampling rate determine frequency range
- Aliasing
- Low-pass filtering

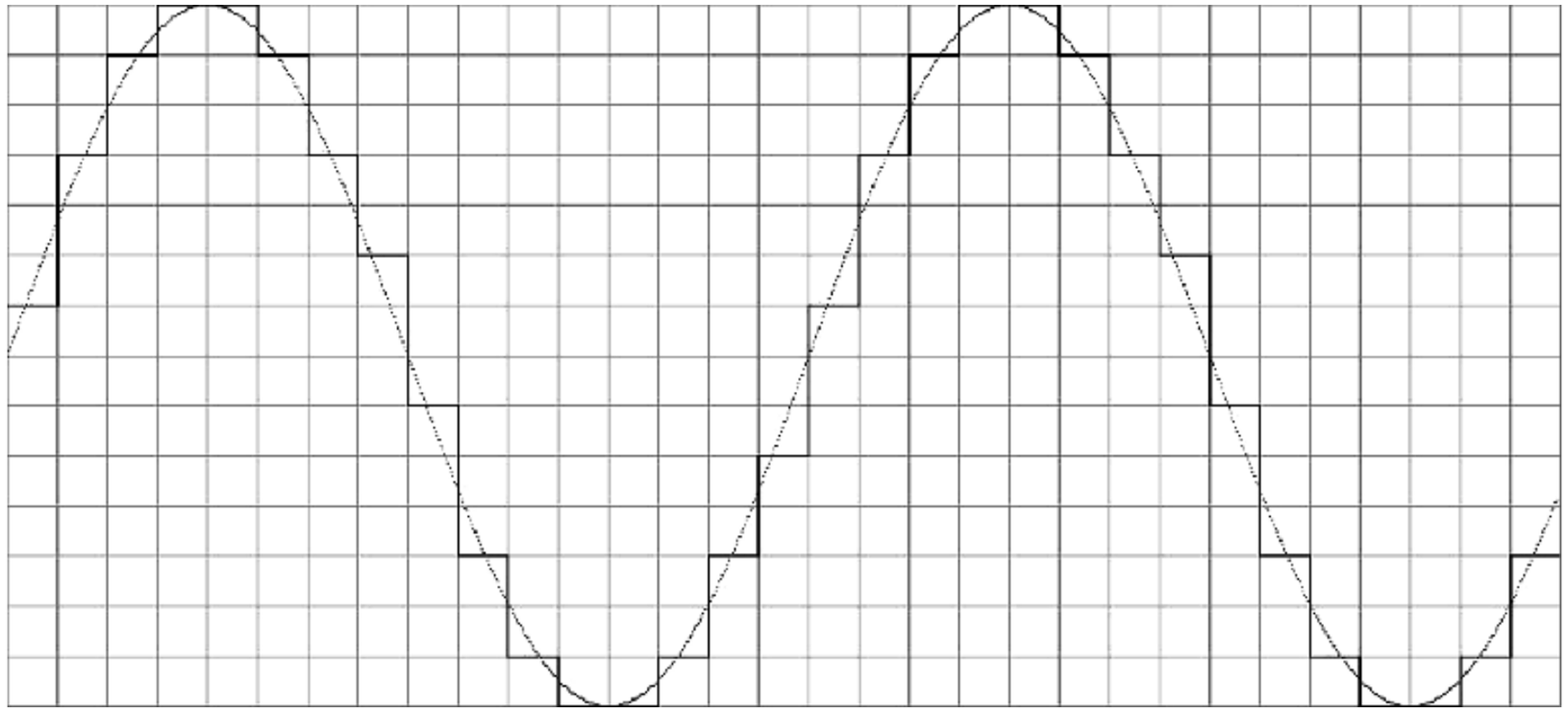


Figure 1.6 Continuous function sampled onto an underlying grid. Note that the values at each sampling step have been quantized to the nearest grid line on the y-axis to give the step function.

The digital sine

For a digital system, it is only possible to deal with discrete samples of time n; with a sampling rate of R per second, time passes such that sample n occurs after n/R seconds. For a sine, the distance traversed in the angular domain of the function input is time $*$ angular velocity $= \omega n/R$, and we can formulate a general sinusoid as

$$\sin(\omega n/R + \phi) \tag{1.4}$$

1.2.6 Filters

- Filter = a device that changes the frequency and/or phase features of a sound
- Filters may strengthen or weaken (and even remove) parts of the sound
- Filters in digital audio basically operate by adding delayed copies of parts of the sound to the present sound
- Filter response = how a filter affects a sound
- Filters at work in 'natural' sound production, e.g. in the human voice, in musical instruments, and in rooms/open spaces
- A spectrogram is a kind of filter bank

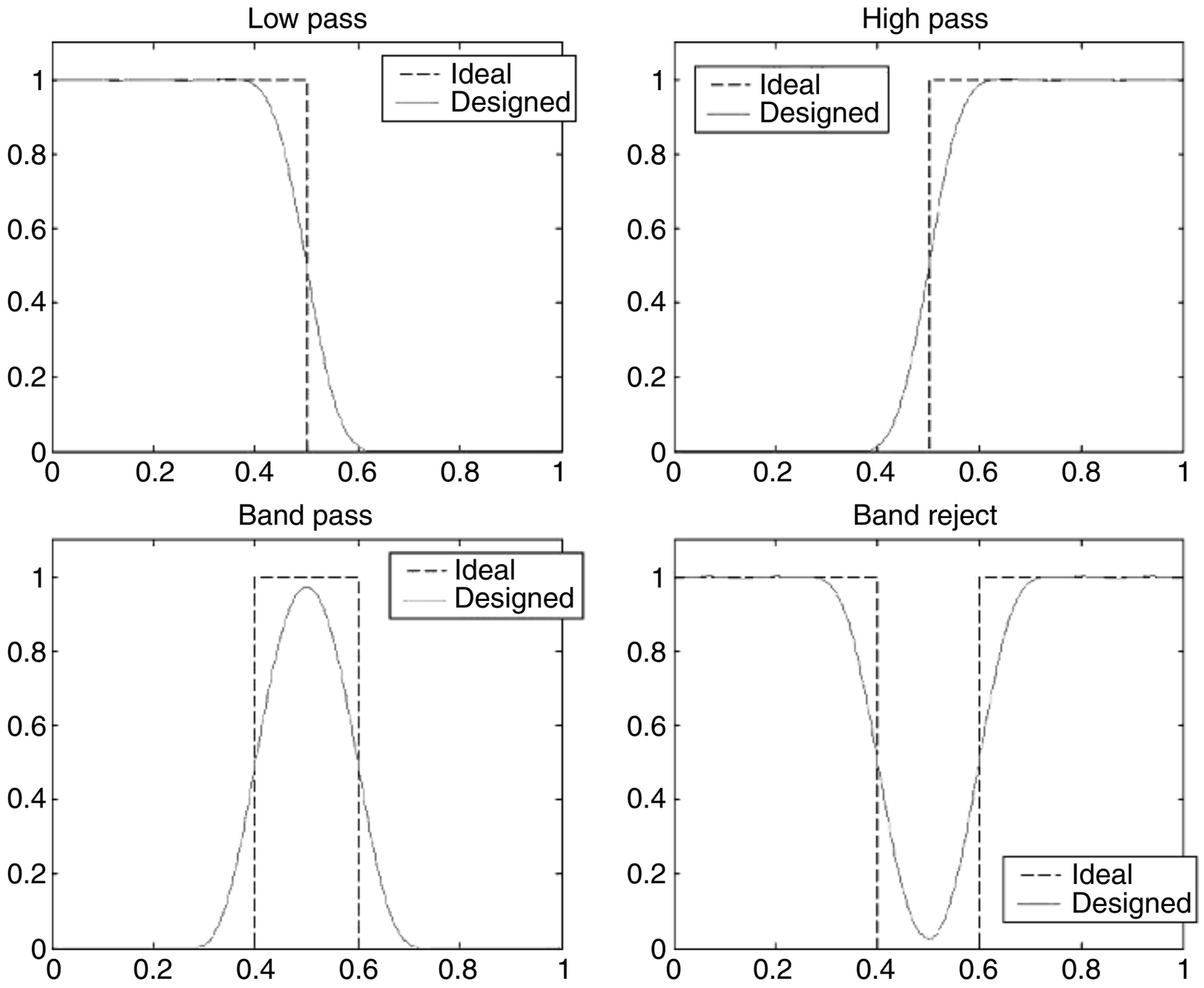


Figure 1.7 Four basic filter types. Magnitude spectra illustrate the frequency responses for four standard filter types. The band-reject filter is also often called a band-stop or notch filter. Each plot shows the frequency response for an ideal filter, and the actual frequency response for some real finite impulse response filters approximating the ideal specification. In this diagram, the x-axis is marked in terms of 'normalized frequency', with a reference of the Nyquist rate, so 0 = 0 Hz and 1 = half the sampling rate. This makes the diagram independent of the sampling rate.

1.2.7 Timbre

- Sometimes defined as that which distinguishes one sound source from another
- Timbre = multidimensional phenomenon
- Basic components:
- Spectrum, usually time-varying
- Dynamical envelope
- Pitch envelope
- Various transients
- Research methods: mapping our various feature dimensions and multidimensional perceptual studies
- Historically, increasing timbre focus in Western music during last hundred years

1.2.8 Space

- Spatial features receiving increasing interest
- Room acoustics obviously part of Western music history (e.g. the size of ensembles, the assumption of reverberation time, etc.)
- Research on psychoacoustic spatial cues in music
- Research on reverberant features of rooms, e.g. by so-called impulse response measurements
- Possibilities of simulations of various rooms
- Link to digital filters (more on this later)

1.2.9 Patching, Signal Flow and Unit Generators

- Modeling musical sound in computers as signal flow
- Metaphor of *patch cord* (or patch bay) from analogue musical instruments
- Also *module* metaphor from analogue instruments: the signal i passed from module to module for various kinds of processing
- Unit generator: generates waveform or noise
- Audio rate signals
- Control rate signals
- Advantage of module and signal flow pictures: see what happens
- More on this with Max-msp or PD

1.2.10 Computer Music Software

- Huge amount of music software available
- Commercial
- Non-commercial
- Ready-made
- Flexible (toolkit-type)
- Comment: for research on musical sound ('Lydanalyse'), consider both good analytical software such as SPEAR, SonicVisualiser, MIRtoolbox, etc., and generative software such as PD, Max-msp, etc. for analysis-by-synthesis approaches

1.2.11 Programming and Computer Music

- Balance between ready-made and doing things from scratch
- Suggested compromises in our present course
- Useful software requiring knowledge of other programing, e.g. *Modalys* (physical model synthesis) requiring *Lisp*
- Also long learning process for 'semi-ready' software like Max-msp (e.g. going into matrix programing in Jitter)
- Collins' use of pseudocode: a general scheme of how to formalize some process

1.2.12 Representing Music on a Computer

- Ideally all the different levels from the sub-symbolic (signal-level) to the symbolic (notation or other representation) and the supra-symbolic (chunk-level of grooves, patterns, etc.) should be interrelated
- The 'signal-to-symbol' problem
- Discussions of timescales
- Arguments for the importance of the sonic object or chunk-level representations
- Much attention to different timescales in connection with MIRtoolbox work later

1.3 A Whirlwind History of Computer Music

- We should also think of categorization and discretization in Western music history as part of the story
- Why has music been so early and easily digitized?
- Analog electronic instruments and the early history of EA music
- Computer music in the 1960s to 1980s
- The explosion of accessible digital technologies from the 1990s and onwards

Chapter 2, Recording

2.1 Recording: A History

- Interesting, and now, by music history standards, fairly long history of recording
- From Edison wax rolls to digital recording by way of shellac discs, wire recorders and tape recorders
- But also history of formats for public diffusion
- And: history of source control recordings from early mechanical instruments to MIDI
- As well as a history of digital file formats for computers
- High-quality recording and processing technology now readily available ('democratization' of music technology)

2.2 Musical Instrument Digital Interface (MIDI)

- MIDI = basically a symbol-level source generation coding scheme (as opposed to signal coding schemes)
- Origin in the early 1980s (based on 1980 technology) and meant for communication between electronic instruments (synthesizers, drum machines)
- Based on Western musical notions of pitch (without enharmonic spelling), but with added control data possibilities (some degree of mapping freedom)
- Serial and slow 31250 bits/second, could lead to choking, and various other shortcomings

2.2 Musical Instrument Digital Interface (MIDI)

- In spite of its shortcomings, MIDI has persisted and been continued in much faster transmission systems
- As a symbol system, has also been useful for computer assisted musicology, e.g. Max, Open Music, MIDItoolbox and various other research-related MIDI software
- Also been extended into much improved notation software
- In general, interesting also as a project for formalizing musical features, cf. an overview of MIDI controllers

2.3 Virtual Studios

- Readily available high speed computers made previously high-end studio technology fit into PCs
- Digital audio workstation (DAW) usually integrates recording, storing, processing, and playback facilities, hence 'virtual studios', see list of usual software functions in table 2.3
- However, challenges of interface ergonomics calls for various physical/gestural controllers that can handle multidimensional input
- Again: be careful of closed, proprietary formats, as these may quickly become inaccessible

2.4 File Formats and Audio Codecs

- Different formats for coding digital audio
- Again, be aware of format incompatibilities
- Lossless:
- Most used: AIFF (and WAV), 16 bit resolution and 44100 khz sampling rate
- Other open and often used formats with higher bit resolution and sampling rates (both for recording and for processing)
- Lossy:
- Various schemes for data-reduction based on removing elements that are considered perceptually redundant, e.g. the so-called MP3 format

2.5 Spatialization

- Spatialization = sound in space
- Localization
- Surround
- Binaural
- Various modes of playback
- Spatialization involves very many features of sound, and these can be measured and studied by various signal models such as delay lines and filters
- Spatialization involves many features of hearing, and these can also be experimentally studied

2.5.1 Spatial Hearing and Room Acoustics Primer

- Distance: inverse square law, properties of reverberation and atmospheric effects
- Location: interaural time difference, interaural intensity difference, body filtering, law of first wavefront, head movement
- Motion: Doppler effect, parallax effect (closer objects moves faster across field of hearing)
- Room acoustics: early reflections, decay of diffuse reflections, interaural decorrelation
- Spherical coordinate system where listeners place sound sources in 3-D
- Learning of source features important for spatial sound perception

2.5.2 Surround Sound Configurations and Multichannel Formats

- Number of speakers and their placement
- Possibility to create artificial spatial sensations
- Overview of spatialization technologies in table 2.6
- Software enables control of source location, various reverb phenomena, filtering, Doppler, etc.
- Diffusion probably one area of much effort in coming years, cf. our own 36 speaker system in the MoCap lab available for your use!

2.6 Recording Tips and Tricks

- Recording a matter of aesthetic choices
- Some pragmatic concerns in relation to sound research:
- Good signal-to-noise ratio
- Avoid distortion, i.e. regulate level
- Take care of your hearing
- Be aware of the room acoustics coloring the recording
- As for mixing, also good idea to focus on clarity and salience of sounds

2.7 Sampling

- Confusing terminology:
- 'Sampling' = discretization of a continuous waveform
- 'Sampling' = recording short sound fragments or taking short sound fragments from other recordings
- Here: 'Sampling' as the art of audio collage
- Various copyright issues involved here (see section 2.7.2), check the FreeSound project
- Historically, musique concrète and sillon fermé the beginning of sampling
- Now in various different genres from Zappa to DJ scratching (see table 2.7)

2.7.3 Sample Playback

- Changing sample rate in playback shifts spectrum
- Better to use a digital vocoder for changing playback speed
- Changing playback by sample decimation (erasing each nth sample), requiring repair by interpolation to avoid noise
- Sampled musical instruments: require multiple samples at different loudness for each pitch
- Sampled instruments incapable of contextual effects (i.e. of coarticulation, more on this later)
- Granular synthesis a popular sample-based method for sound generation

Chapter 3, Analysis

Some considerations of timescales:

Timescales of sonic features and of music-related actions (from Snyder 2000), our focus in the approximately 16 to 0.2 Hz range:

Table 1.1
Three Levels of Musical Experience

	Events per second	Seconds per event
	16,384	1/16,384
EVENT FUSION	8,192	1/8,192
(early processing)	4,096	1/4,096
	2,048	1/2,048
Functional units =	1,024	1/1,024
individual *events* and	512	1/512
boundaries; pitches,	256	1/256
simultaneous intervals,	128	1/128
loudness changes, etc.	64	1/64
	32	1/32
MELODIC and RHYTHMIC GROUPING	16	1/16
(short-term memory)	8	1/8
	4	1/4
Functional units =	2	1/2
patterns; rhythmic and	1	1
melodic groupings,	1/2	2
phrases.	1/4	4
	1/8	8
	1/16	16
FORM	1/32	32
(long-term memory)	1/64	1 min 4 sec
Functional units =	1/128	2 min 8 sec
large scale *constancies*;	1/256	4 min 16 sec
sections, movements,	1/512	8 min 32 sec
entire pieces.	1/1,024	17 min 4 sec
	1/2,048	34 min 8 sec
	1/4,096	1 hr 8 min 16 sec.

Timescales in music perception

- Various different thresholds for the perception of pitch, timbre, simultaneity, order, timbre, etc. (e.g. Moore 1995)
- Thresholds for perceiving various rhythmical, textural, melodic, harmonic, etc. patterns: (we assume) as long as they take to unfold
- Thresholds for recognition rather different, cf. Gjerdingen & Perrott 2008 indicating recognition down to 250 ms

Timescales in music perception

- In music-related action, often various simultaneous different temporal layers corresponding to layers in sonic textures, i.e. hierarchies of timescales
- Also different simultaneous layers in sound-accompanying movements, e.g. dance, studied by FFT, EMD, and other methods
- And: so-called *phase transition* thresholds in action, e.g. between singular strokes and tremolo (and other actions)
- Inspired by Schaeffer 1966, we work according to the following conceptual three-level model:

Conceptual three-level model of timescales:

- *Sub-chunk level*: Continuous sound and actions below the chunk level of duration (i.e. below roughly 0.5 seconds)
- *Chunk level*: Holistically perceived fragments of sound and action roughly in the 0.5 to 5 seconds range
- *Supra-chunk level*: Concatenations of chunks into larger scale units, i.e. into sections, movements, and whole works.

3.1 Sound Analysis

- Two-fold aim of sound analysis:
- Find out features of the signal from low-level to high-level
- Find out features of listening and subjective experience of musical sound
- And: correlate signal features with subjective notions
- But: not all signal features intuitively relevant for subjective sensations, e.g. large-scale statistical surveys in MIR
- And: not always easy to pinpoint the basis for our subjective experiences in the signal

Fourier analysis: examining a signal with probe signals:

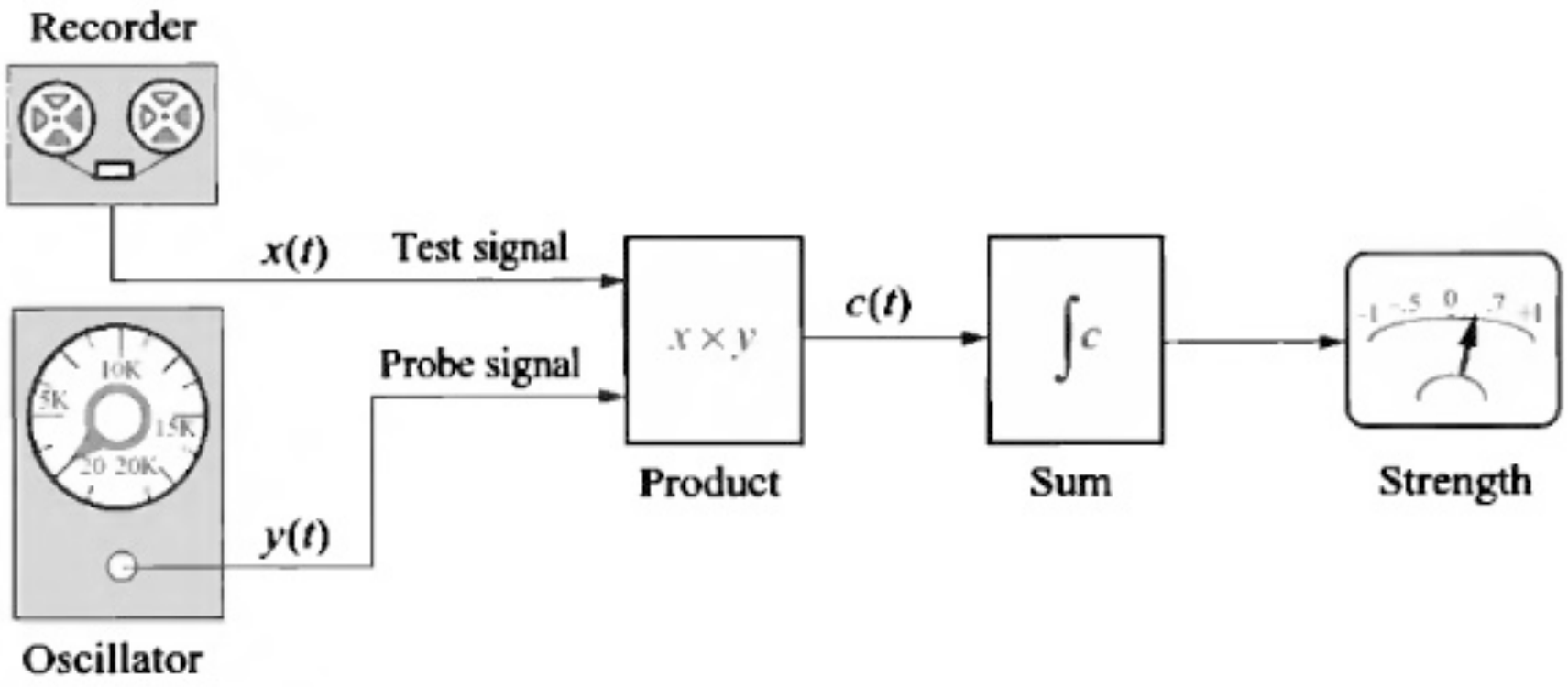


Figure 3.6
Frequency analyzer.

Fourier analysis and time-frequency uncertainty:

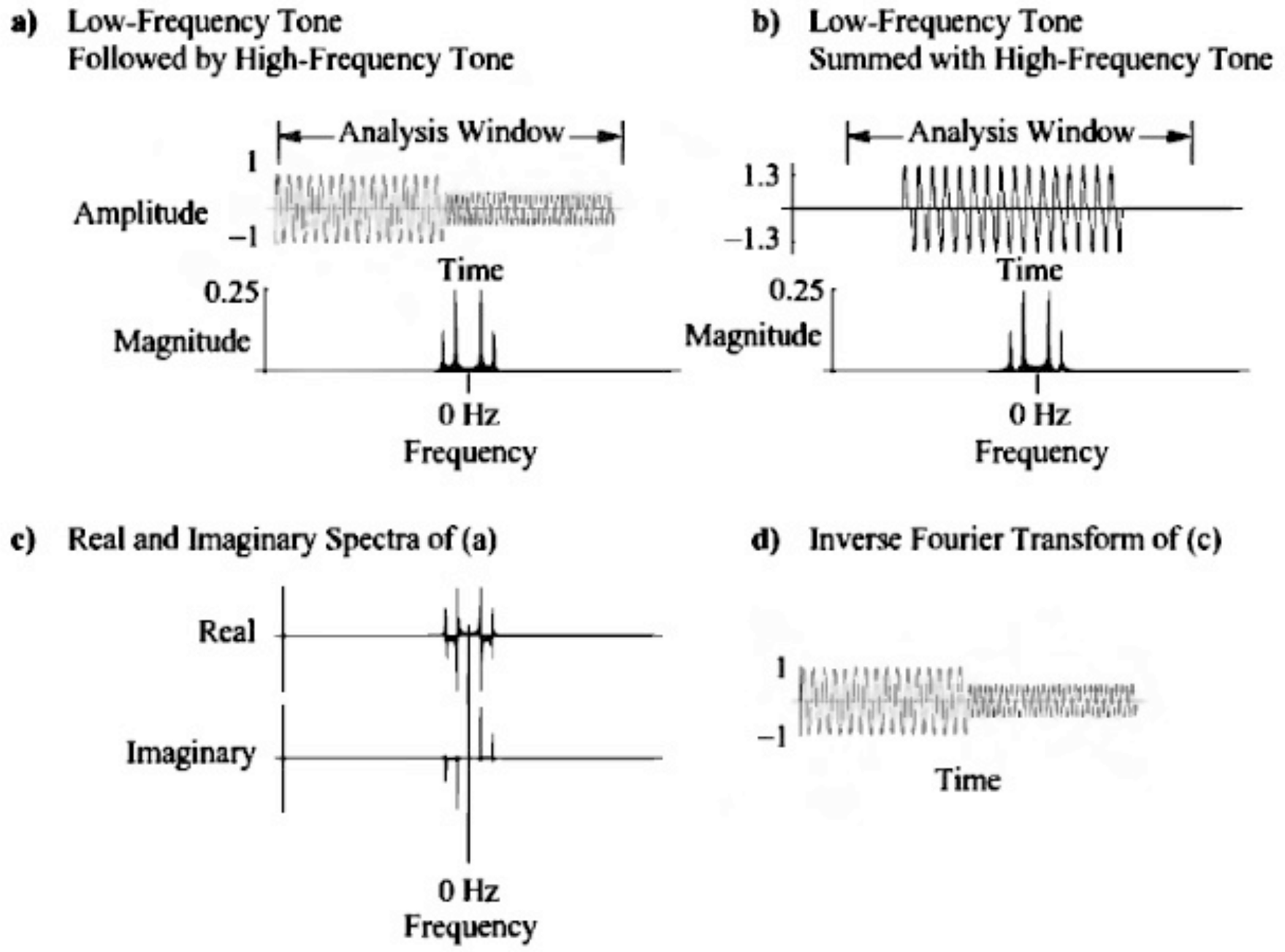


Figure 10.1
Magnitude fourier transform of two signals.

Gabor's strategy for reducing the time-frequency uncertainty:

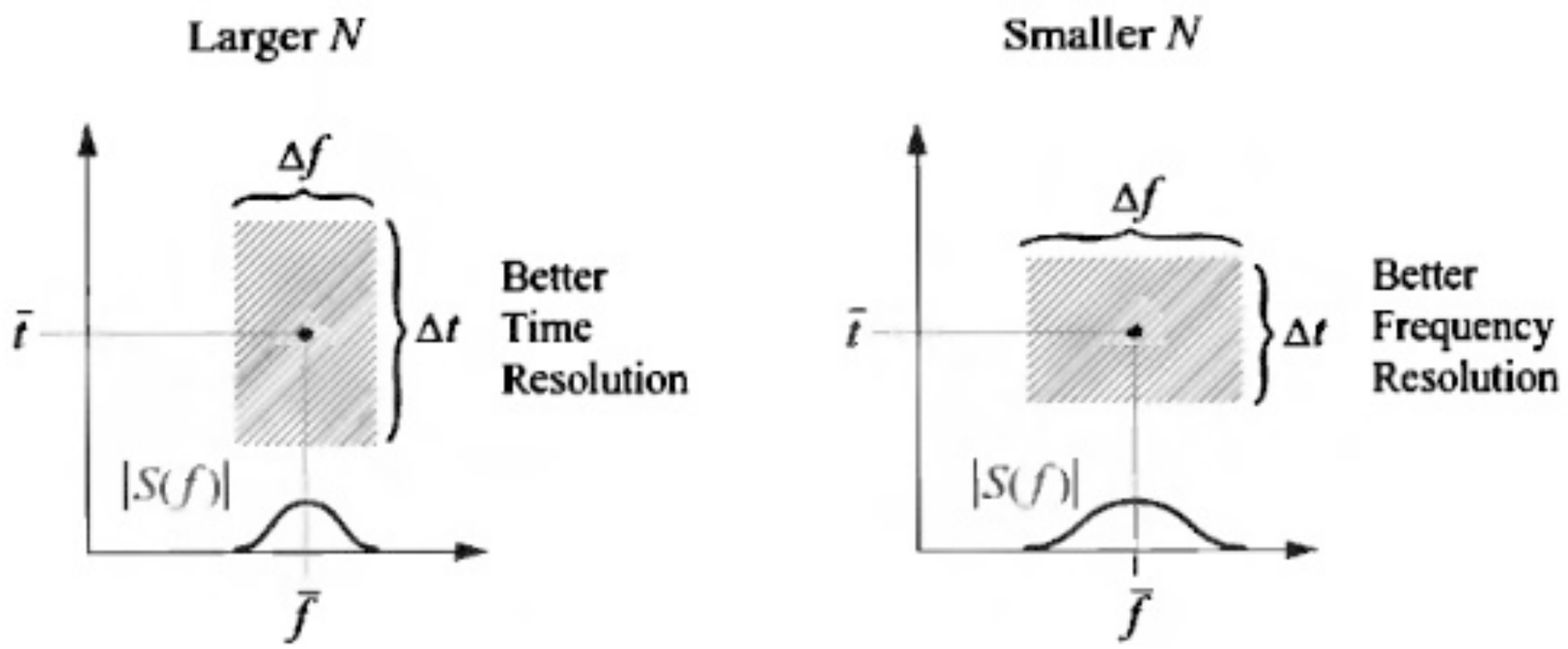


Figure 10.5
Heisenberg boxes.

3.2 Fourier Analysis and the Phase Vocoder

- Our basis: discrete representations of signals. Good idea to reflect on issues of 'in time' and 'outside time'
- The discrete Fourier transform (DFT) basis here
- Most natural sounds vary quite a lot in time, hence strategy here to take a number of 'snapshots' to see what happens moment-by-moment
    - Each snapshot = a window containing N-number of sample points
    - The FFT algorithm requires sample number be in the power of 2 to work fast ('zero padding' an alternative)
    - STFT as a sequence of DFTs, to be used for various representations and further processing and resynthesis
    - 'Hop size' indicate degree of overlap of windows

3.2.1 Short-time Fourier Transform

- Fourier transform assumes infinitely long and perfectly harmonic sounds
- Hence: problems with shorter, time-varying and inharmonic sound
- STFT in general: a compromise between time and frequency, i.e. good frequency resolution vs. good temporal resolution ('time-frequency uncertainty')
- STFT works by windowing and then multiplying the waveform with equally spaced partials (as many as there are sample points in the windowed waveform)
- The result of these multiplications indicate presence of various partials, hence, results in a spectrum

3.2.1 Short-time Fourier Transform

- Understand the STFT (and Fourier transform in general) as asking questions about what frequencies do we have in a sound, and relate this to additive synthesis where you add sinusoids to create a sound
- The STFT uses complex numbers, i.e. numbers that can contain information about frequency, amplitude and phase, in probing the windowed waveform
- Phase information important for resynthesis of the waveform from the spectrum, and for various processing, e.g. the digital phase vocoder (SPEAR)
- Window length decisive for frequency resolution
- Window shape decisive for artefacts of smearing, see figure 3.1

3.2.2 Further Refinements to Fourier Analysis

- Problems with mismatch of frequencies in the waveform and the analysis grid
- Experiment with different window sizes and hop sizes (as well as window shapes) to get better results
- Also alternative methods (more on this later)
  - Tracking spectral peaks can give important information of spectral envelope and formants
  - Peak matching by the phase vocoder tries to follow sinusoids, their 'births' and 'deaths' (by specifying certain thresholds), resulting in a fairly clear partials representation that may be manipulated in various ways. So again: play with SPEAR!

3.2.3 Resynthesis after Sinusoidal Modeling

- Extracted sinusoidal information can be used in resynthesis, similar to additive synthesis, often called 'oscillator bank' (see 'summing sinusoids')
- IFFT is the other resynthesis model, converting spectral frames back to a waveform
- IFFT requires both magnitude (i.e. amplitude) and phase (i.e. precise temporal information) data to reconstruct the original waveform without too much distortion.
- Also here, various overlap options to improve the result

3.2.4 Not All Sounds are Periodic: Coping with Noise

- Subtracting sinusoid-based resynthesis sound from the original sound, we are left with a noise residue
- Hence: deterministic + stochastic decomposition of musical sound, fits nicely with the mix of order and chaos in most natural sounds
    - Deterministic (regular, predictable) and stochastic (irregular, chaotic) components both in the quasi-stationary part of a signal (e.g. the white noise components of a sustained violin tone) and in the transients (attack part and other fluctuations)
    - Much 'liveliness' in sound due to this mix of order and chaos, hence challenge for Lydanalyse here!

3.3 Alternative Representations for Analyzing Sound

- Basic problem here (also of 'order vs. chaos') is the time-frequency uncertainty
- Dennis Gabor: time-frequency quanta, i.e. instead of long waveforms at different frequencies, short quanta of sound that could be combined
- Multi-resolution analysis: having different window sizes overlap (figure 3.3) to capture both 'fast' and 'slow' components
- Much work on so-called wavelets in past decades, slowly providing better means for detail analysis of sound

3.4 Auditory Models

- What we can find out by studying the signal does not always match what we subjectively experience, what Schaeffer called *anamorphosis* (or 'warping') hence:
- Auditory models in sound research have received much attention in past decades
  - These models may encompass the entire range from the signal, by way of our hearing apparatus, up to high level sonic features in our minds
  - Important to recognize the holistic nature of music perception, i.e. that there is a two-way flow (afferent and efferent) of information in perception in general

3.4.1 Physiologically Inspired Models of the Auditory System

- Models of the ear, including components from the outer ear to the auditory-related areas of the brain
- Simulations of various signal processing that seem to go on in the ear, see table 3.1, resulting in alternative representations, e.g. the neural activity patterns or so-called cochleagrams, cf. figure 3.5
- Signal processing based on auditory models possible with various Matlab-based toolboxes, cf. p. 96

3.4.2 Computational Auditory Scene Analysis

- Classical Auditory Scene Analysis as summarized in Bregman 1990: how our listening capabilities have evolved and how sophisticated our capacity for analyzing very complex auditory scenes
- Computational Auditory Scene Analysis (CASA) tries to simulate our capabilities
- Major challenge: source separation, in particular in view of the complex mix of partials in the spectrum at any given time
- General conclusion: we use very much previous knowledge in auditory perception, and advances in CASA and music transcription dependent on this

3.4.3 Perceptual Audio Coding

- Studies of human audition has revealed a number of non-linearities and other discrepancies between sound and our perception of sound:
- Non-linear frequency response
- Critical bands
- Frequential masking
- Temporal masking
- These 'weaknesses' of our hearing has been exploited in audio coding, i.e. compression based on discarding elements we don't hear anyway
- Object coding, a bit more in the future

3.5 Feature Extraction

- Challenge: extract perceptually salient features from the signal, i.e. to establish correlations between our subjective experience and sound
- Feature extraction based on various representations of the sound, both in the time and frequency domains and by auditory modeling
- A number of features have been singled out, cf. table 3.3, but of course also possible (and desirable) to single out new ones
- Combined bottom-up and top-down approaches here
- Intense work on feature extraction in MIR, but not all features intuitively perceptually salient
- Later, much work with the MIRtoolbox

Table 3.3 Examples of low-level features.

Feature	Description	Calculation
ZCR	Count (positive) zero crossings within N samples	$\sum_{k=0}^{N-2} x(k+1) \geq 0 \wedge x(k) < 0$
RMS	Root mean square amplitude calculated over N samples	$\sqrt{\frac{\sum_{k=0}^{N-1} x(k)^2}{N}}$
Max power	Maximum power in a block of N samples; often used in sample editor waveform displays when zoomed out	$\max_{k=0}^{N-1} x(k)^2$
Spectral centroid	Statistical measure over the spectrum	$\frac{\sum_{k=0}^{N/2-1} k\lvert X_m(k)\rvert^2}{\max(\sum_{k=0}^{N/2-1} \lvert X_m(k)\rvert^2, 1)}$
Spectral flux	Change of spectrum between frames	$\sum_{k=0}^{N/2-1} \lvert\lvert X_{m+1}(k)\rvert^2 - \lvert X_m(k)\rvert^2\rvert$
Spectral fall-off	The spectral envelope can be modeled by fitting a curve to the magnitude spectrum. Spectral fall-off fits a single line to model the typical drop in energy at higher frequencies in sound, as one helpful timbral indicator, but more complex models are available	Rodet and Schwarz [2007]
LPC coefficients	Linear predictive coding models the spectrum of the input with a source-filter model; it is a useful compression technique	Gold and Morgan [2000]; Rabiner and Juang [1993]; Makhoul [1975]
MFCCs	Mel-frequency cepstral coefficients; given a spectrum, the **cepstrum** approximates the principal components, and is a useful timbre descriptor; it also deconvolves (separates) an excitation and body response and gives some idea of pitch	Gold and Morgan [2000]; Logan [2000]; Roads [1996, pp. 514–8]

Table 3.4 Examples of higher-level features.

Feature	Description	Review references
Onset detection	Identifying the physical beginning of sound events	Bello *et al.* [2004]; Collins [2005]; Dixon [2006]
Pitch detection (monophonic)	Finding the fundamental frequency that would be selected by the human auditory system	de Cheveigné [2006]; Gómez *et al.* [2003]
Melody extraction	Transcription of a lead melody line, for example, as a sequence of discrete notes	Gómez *et al.* [2003]
Pitch detection (polyphonic)	The more general case of multiple simultaneous voices	Klapuri [2004]; de Cheveigné [2006]; Klapuri and Davy [2006]
Key and chord recognition	Detection of harmony	Gómez [2006]
Beat tracking and rhythm extraction	Determination of tempo, of beat locations and other metrical structure, and of rhythmic patterns	Gouyon and Meudic [2003]; Gouyon and Dixon [2005]
Instrument recognition	Timbral categorization	Herrera-Boyer *et al.* [2003]; Klapuri and Davy [2006]

3.5.1 Pitch Detection

- Pitch is a psychoacoustic feature, not necessarily equal to frequency
- Generally accepted that pitch is an emergent feature dependent on several different features, in particular on the spectrum
  - Complex tones vs. pure tones
  - Periodicity
  - Autocorrelation, cf. figure 3.7
  - Missing fundamental
  - Inharmonic sounds
  - Try out pitch detection in Max-msp or PD!

3.5.2 Onset Detection

- Onset detection likewise sometimes easy, sometimes quite tricky
- The simple version: peaks in the signal
- The tricky version: several competing peaks in the signal
- In general: best results with a combination of time domain and frequency domain features, cf. figure 3.8
- Also issues of subjective sensations of onsets, the so-called 'p-centers' problem

3.5.3 Beat Tracking

- Beat tracking: an even more challenging task
- Subjectively, beat is very often (but not always) quite obvious for listeners
- Beat is generally considered a mental phenomenon, partly due to the signal and partly due to subjective expectations, hence, human capacities for *anticipation* and past learning
- Various models for beat tracking (and meter tracking): cross-correlation and IOI (inter-onset interval histogramming)
- Some examples in Jehan's work, http://web.media.mit.edu/~tristan/Blog/Beat_Tracking_v1.html

3.6 The Transcription Problem

- Given the challenges of CASA and source separation, still a very long way to go for automatic transcription of polyphonic music
- Monophonic music more tractable in terms of pitch, but rhythm still remains tricky
- Transcription also a problem of our Western music theory concepts of discretized pitches and durations
- Other feature extractions may perhaps be more useful for sound research, i.e. more in the direction of Schaeffer-like sonic object descriptions
- Also interesting to focus on global feature extraction, referred to as 'auditory gist perception'
- Some examples: http://www.cs.tut.fi/sgn/arg/matti/demos/polytrans.html

3.7 Machine Listening and Causal Realtime Analysis

- Realtime: generally understood as that which happens so fast that we believe it is instantaneous, whereas in fact, there are latencies everywhere in music technology (and in music as such)
- Causal: that which is only based on what has been presented recently, e.g. as in causal filters, and 'non-causal' means that which is outside time, e.g. processing a whole audio file before we listen
- Realtime processing in interactive music, e.g. NIME, a fast growing field
- Anticipation one of the most intriguing elements in human perception and action, probably essential for understanding our experience of music (cf. our work on music-related movement)

Chapter 4, Processing

4.1 More on Signals

- Traditional distinction between synthesis and processing, but in practice, a fusion of these
- Digital Signal Processing, DSP, a vast field of research and concerning most scientific and technological domains. Music-related DSP draws on this vast field
- Proliferation of music-related DSP with very many modules in music software (and hardware)
- Challenge of evaluating perceptual and aesthetic effects of DSP models

4.1.1 Signals and Sampling Rates

- Changing sample rates for signals: resampling
- Upsampling and downsampling will have significant consequences for sonic features
- Will need various processing to avoid undesirable results, e.g. interpolation between samples and low-pass filtering to reduce noise
- Normalization = to bring different signals up to nominally equivalent levels, e.g. try this in Audacity when you have edited various sound excerpts

4.1.2 Ring and Amplitude Modulation

- Modulation = general term for changing a signal by some other input, e.g. in the sub-audio range (<20hz) by making a vibrato, tremolo, etc., or in the audio range (>20hz) resulting in timbral and other effects
- Sidebands = new frequencies introduced by the modulation
- Carrier (C) and Modulator (M)
- Ring modulation = sum + difference of C and M
- Amplitude modulation = sum + difference of C and M, as well as C
- Try this with both sine tones and complex tones!

4.1.3 Mixing and Splicing

- Mixing signals: general principles of having enough headroom and avoiding overload
- Envelopes for gain control: breakpoints and functions (various exponents)
- Splicing to avoid discontinuities in the signal
  - Various curves for crossfades
  - Try experimenting with different curves and evaluate the smoothness of the transitions between sound fragments

4.1.4 Delays

- Delay = general principle in both room acoustics and in musical instruments
- Delay lines: allows us to experiment with different reverberations and also making physical models of musical instruments
- Summing delayed (and attenuated) copies of signals essential for understanding filters, rooms, and instruments
- Variables: delay length, attenuation, phase change, and tappings
- Very instructive to experiment with delay lines!

4.2 Convolution and Filters

- Convolution = the operation of point-by-point summing of delayed and multiplied copies of one signal by another signal

Here is a concrete example of convolution.

$$h(n) = f(\cdot) * g(\cdot) = \{1, 2, 3, 4, 5\} * \{1, 2, 3\}. \tag{4.4}$$

	1	2	3				
		2	4	6			
			3	6	9		
				4	8	12	
					5	10	15
$h(n) =$	1	4	10	16	22	22	15

4.2 Convolution and Filters

- Note that convolution is commutative:

$$h(n) = g(\cdot) * f(\cdot) = \{1, 2, 3\} * \{1, 2, 3, 4, 5\}. \tag{4.5}$$

	1	2	3	4	5		
		2	4	6	8	10	
			3	6	9	12	15
$h(n) =$	1	4	10	16	22	22	15

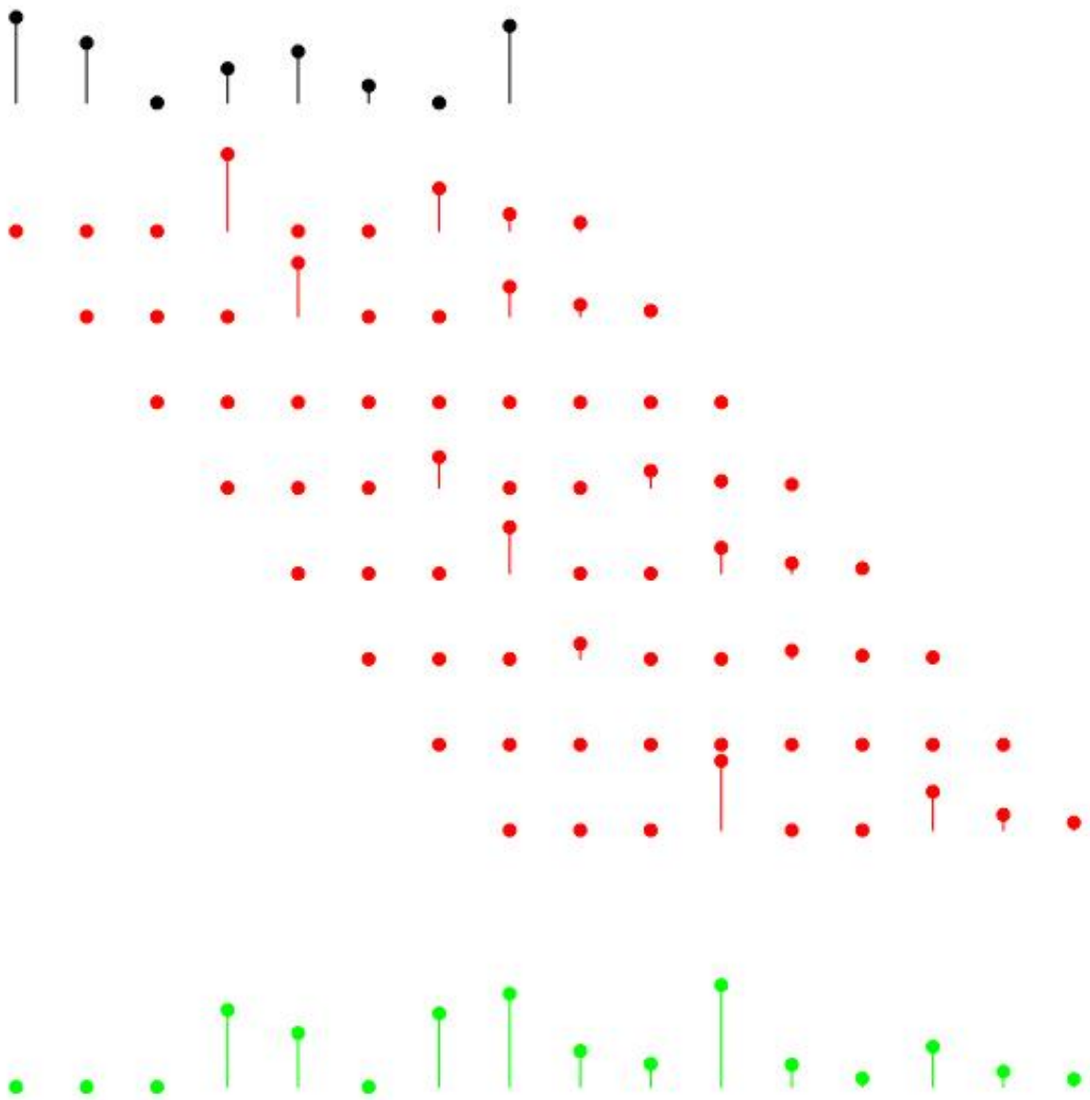
Notice that equations (4.4) and (4.5) produce the same result, indicating that

Convolution is commutative.

- Convolution = one of the most basic operations in digital audio

4.2 Convolution and Filters

- Impulse response is 0,0,0,0.45,0,0,0.25,0.1,0.05
- Input signal is 1,0.7,0,0.4,0.6,0.2,0,0.9 (along the top)



4.2.2 Convolution and Multiplication

- Multiplication of two signals in the time domain convolves them in the frequency domain
- Convolution of two signals in the time domain multiplies them in the frequency domain
- Hence, an important symmetry here

4.2.3 More on Filters

- Impulse response = how a room or other resonating body, e.g. a musical instrument, reacts to an impulse
- Two main classes of filters:
- FIR = Finite Impulse Response combines delayed (and usually attenuated) samples to the present
- IIR = Infinite Impulse Response introduces feedback into the present
- The unit circle with r<1, r=1, or r>1 and the consequences of this
- Frequency response
- Phase response

4.2.4 Examples: Comb and All-pass Filters

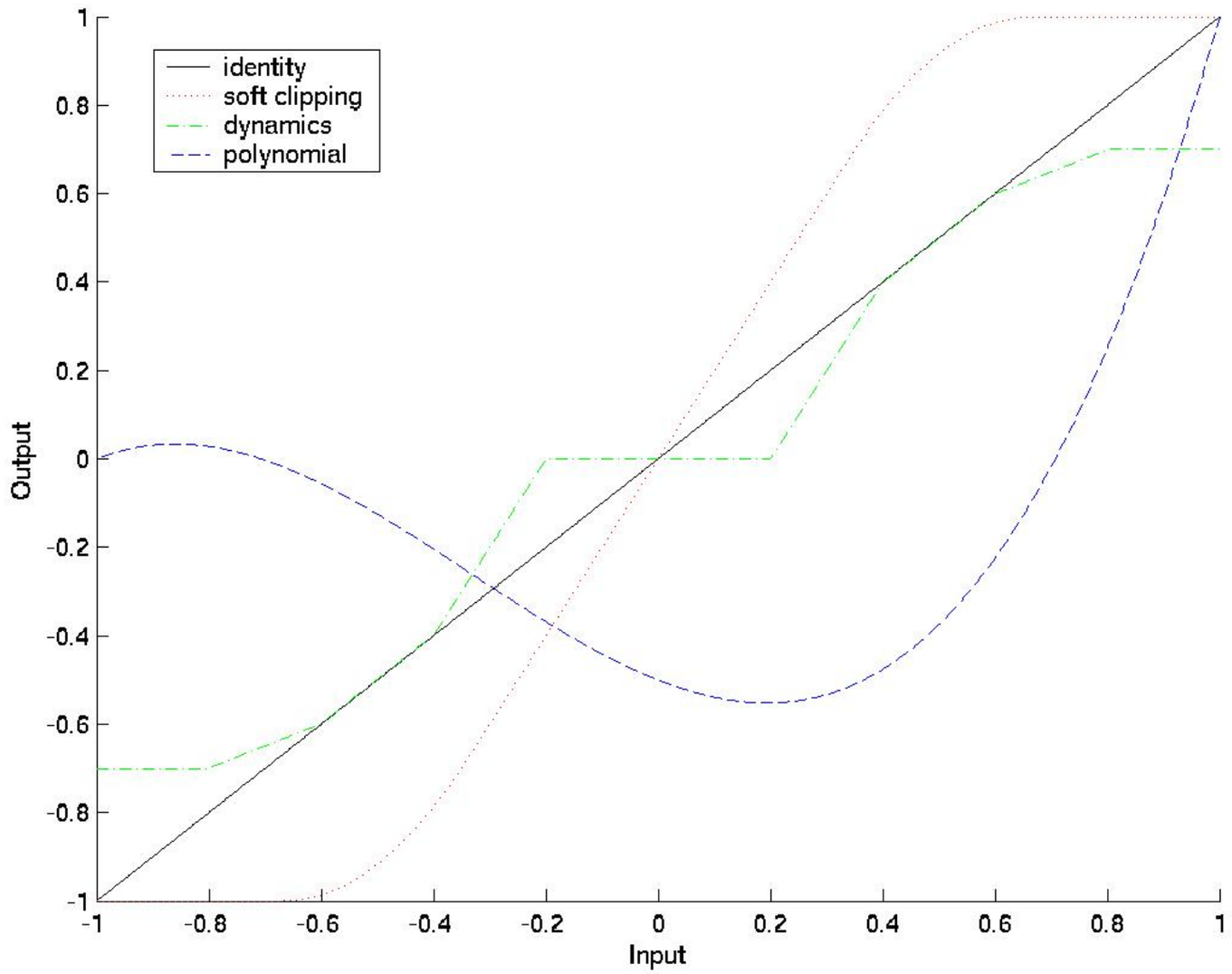
- Comb filters: dips in the spectrum (hence the name)
- Comb filters very useful in studying features of rooms and instruments
- All-pass filters: flat frequency response but varying phase response, hence, may result in dips and peaks because of phase cancellations or phase reinforcements

4.3 A Compendium of Marvelous Digital Audio Effects

- See overview in table 4.2
- Stay updated with the DAFx conferences
- Many effects can be combined
- Challenge of more research on the perceptual and aesthetic aspects of effects processing remains!

4.3.1 Dynamics Processing, Distortion and Waveshaping

- Understanding transfer functions:



4.3.2 Time Stretching and Pitch Shifting

- Various tools for time stretching and pitch shifting important for timbral research, in particular for studying transients and formant features
- A family of FFT-based related models here, cf. the digital phase vocoder, and challenge is to avoid artifacts (noise, distortion, etc.) in the output
- Overlap-add (figure 4.11) a basic strategy

4.3.3 Implementing Spatialization

- Spatialization is increasingly becoming part of effects processing
- Various perceptual cues that can be manipulated in the software:
- Source positioning
- Source motion
- Room features
- And to be explored (in our lab!): the effect of positioning of spectral components in space

Chapter 5, Synthesis

5.1 The Space of Sound Synthesis Algorithms

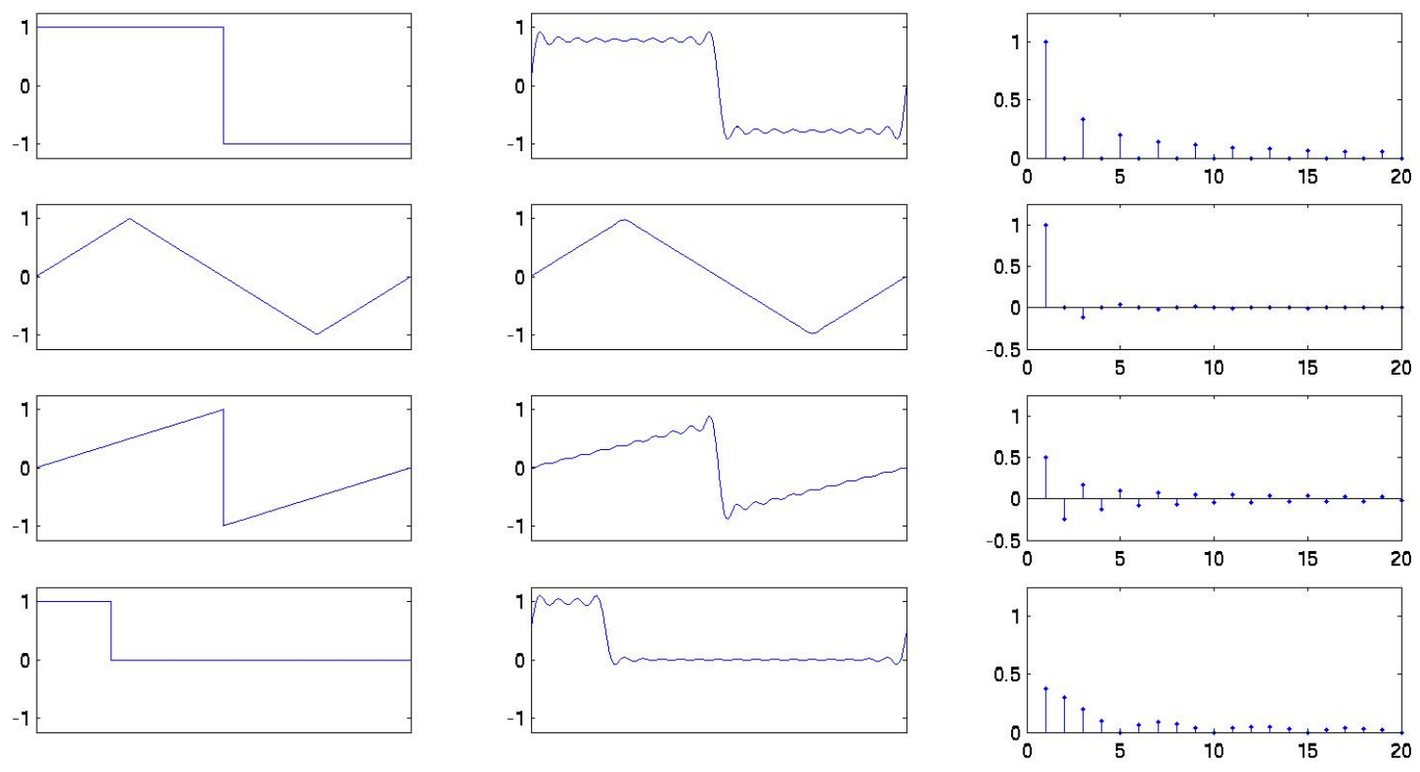
- Main categories of synthesis:
- Processed (variably so) recordings
- Spectral models
- Physical models
- Abstract algorithms
- But also considerations of:
- Sound quality
- Flexibility
- Generality
- Computation cost

5.1.1 Classic Algorithms: overview in table 5.1

- Sample playback (with variants)
- Granular synthesis
- Wavetable synthesis
- AM, FM
- Subtractive (including source-filter models)
- Additive synthesis
- Spectral modeling
- Non-standard modeling
- Physical modeling

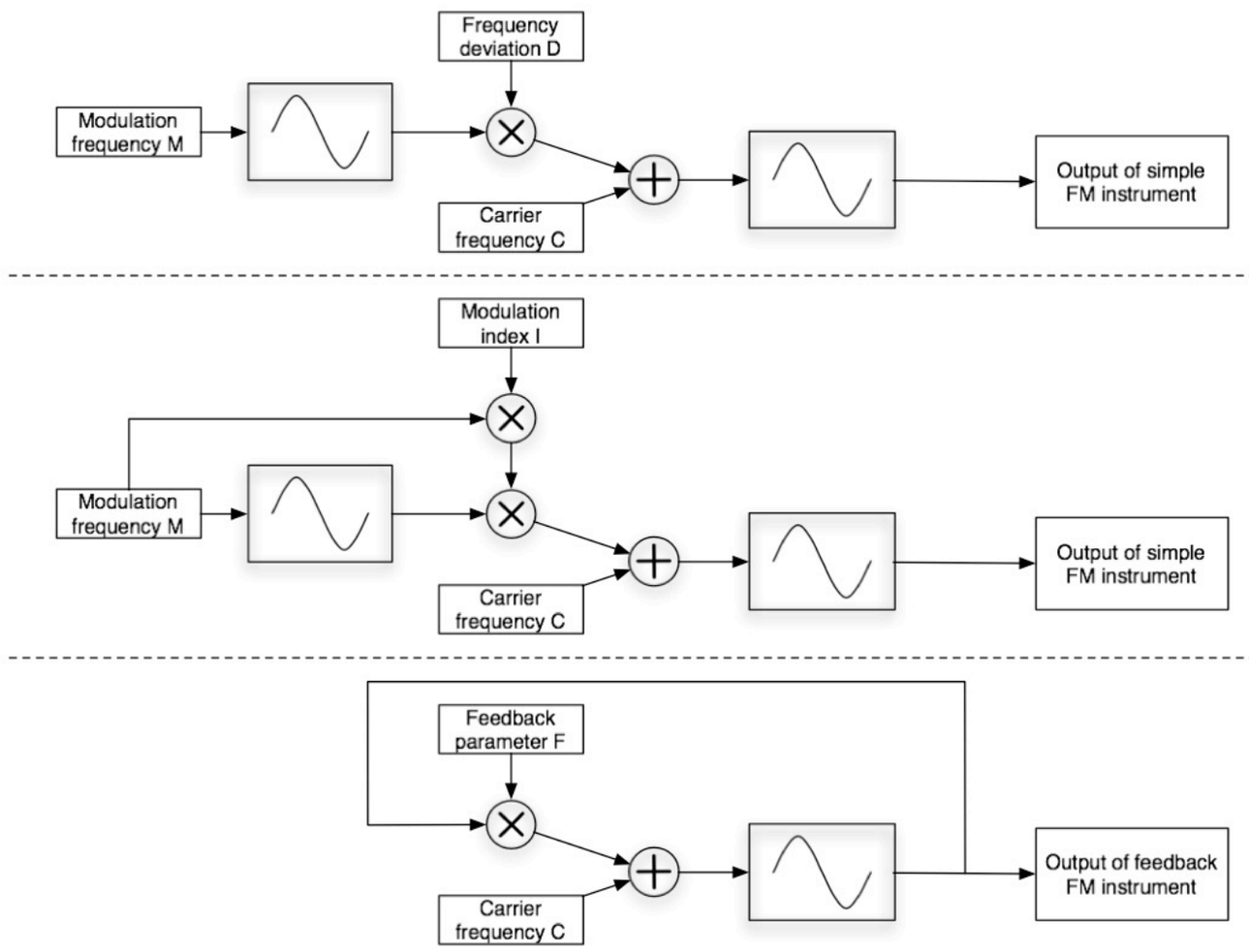
5.1.2 Waveforms and Wavetables

- Analog synthesizers and wave shapes
- Additive production of wave shapes
- Wavetables for efficiency
- Ideal, band-limited, and spectra of some waveforms:



5.1.3 Frequency Modulation

- Very economical in terms of computation
- Frequency modulation in the audio range (>20hz) produces sidebands, and envelopes on C, M, and I produce time-varying spectral features



5.1.4 Granular Synthesis

- Basically, combining very short fragments of sound, either from sound files or from synthesis
- Has interesting relationship to time-frequency issues, e.g. so-called 'Gabor grains' and wavelets
- Challenge of making smooth transitions between grains
- Question of flexibility for different kinds of sonic results
- Various models of concatenative (or diphone) synthesis may be more flexible alternatives

5.1.5 Feature-based Synthesis

- Extracting features from various sounds and using these in synthesis
- Concatenative synthesis can be based on feature extraction and feature combination, e.g. CataRT
- Various results from MIR can help in controlling synthesis
- Genetic algorithms and other trial-and-error strategies for finding essential features of sound in the control of synthesis

5.2 Physical Modeling

- Physical modeling = some kind of imitation of the physical process of sound-production, ranging from the highly simplified (e.g. Karplus-Strong) to the very complex (e.g. finite element method simulations of instruments)
- Previously impractical because of computational cost, now much more feasible
- Ecological advantages: the model will very often behave predictably in terms of sonic results
- Various sub-classes: analytic/numerical (acoustic equations), mass-spring (physics of energy dissipation), modal synthesis (simulating vibration modes, and waveguide synthesis (simulating energy propagation within confines)

5.2.1 Waveguide Synthesis

- The Karplus-Strong model: a burst of white noise repeated in a loop with some filtering
- More sophisticated versions: adding modules to a simple basic scheme in order to bring in more 'life-like' elements, e.g. for energy loss at so-called 'scattering junctions'
- Also possible to make physically impossible instruments, e.g. a "guitar" with the resonance of a tamtam

5.2.2 Singing Voice Synthesis

- Intense research and development in the linguistic community with the goal of having machine-generated speech sound natural
- Many models of singing voice available, both more signal-based and more physical model-based
  - Basic scheme: source (vocal folds) + filter (the entire oral cavity, tongue, lips)
  - Diphone modeling very useful for exploring transitions and coarticulation