EBU5303

Multimedia Fundamentals

Digital Video and Audio

Dr. Marie-Luce Bourguet
marie-luce.bourguet@qmul.ac.uk

Learning Objectives

- Apply the Nyquist theorem to avoid digital audio aliasing.
- Relate quantisation level and dynamic range of an audio file.
- Calculate decibels from air pressure.
- Calculate the signal-to-quantisation noise ratio.
- Interpret the spectral analysis of an audio wave.
- Describe the MIDI format.

Reading

http://burg.cs.wfu.edu/TheScienceOfDigitalMedia/Chapter4/Ch4ScienceOfDigitalMedia.pdf

4.2 Audio Waveforms

4.4 Sampling Rate and Aliasing

4.5.1 Decibels and Dynamic Range

4.6.1 Time and Frequency Domains

4.8 MIDI

http://digitalsoundandmusic.com/

Reading

Fundamentals of Multimedia, by Ze-Nian Li, Mark S. Drew, Jiangchuan Liu (3rd edition)

Chapter 5: Fundamental Concepts in Video

Chapter 6: Basics of Digital Audio

Agenda

- A video is a sequence of images
- A sound is characterised by its frequency (pitch) and amplitude (loudness)
- CD standard quality is 44,100 Hz (sampling) and 16 bits (quantisation)
- Speech signals contain 3 types of sound, some of them are used for speech recognition
- MIDI format for music stores information such as instrument specification, beginning and end of a note, basic frequency, etc.

Video - definitions

- **Video** is the technology of electronically capturing, recording, processing, storing, transmitting, and reconstructing a sequence of still images representing scenes in motion.
- **Frame rate**: the number of still pictures per unit of time of video.

- **Analog video**: video recording method that stores continuous waves of red, green and blue intensities.
- **Digital video**: video recording system that works by using a digital rather than an analog video signal.

Refresh rate and frame rate

- The **refresh rate** is the number of times in a second that the display hardware draws the data (i.e. repeated drawing of identical frames).
- The **frame rate** measures how often a video source can feed an entire frame of new data to a display.
- Typical rates: 24, 25 or 30 frames per second (frame rates) ; 60, 75 or 120 Hz (refresh rates).

Frame Rates

Video Type	Frames Per Second (fps)
NTSC	29.97
PAL	25
SECAM	25
Motion Picture Film	24

NTSC was 30 fps for black-and-white TV, Frame rate was lowered to 29.97 fps to accommodate for color encoding.

Interlaced vs Progressive

- **Interlaced** scanning displays alternating sets of lines. Because each field happens so quickly we are given the illusion of a whole image.
- **Progressive** video displays the entire image.

Interlaced

Progressive Scan
(Non-interlaced)

Exercise

A 30fps digital video uses 352 by 255 pixels video frames with a pixel depth of 8.

i) Calculate the size of 1 second of data.

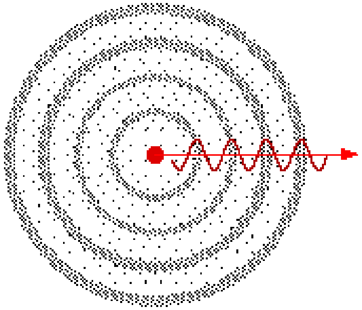
ii) What compression ratio would be needed to transmit 1 second of data in real-time over a 64 Kbps communication channel?

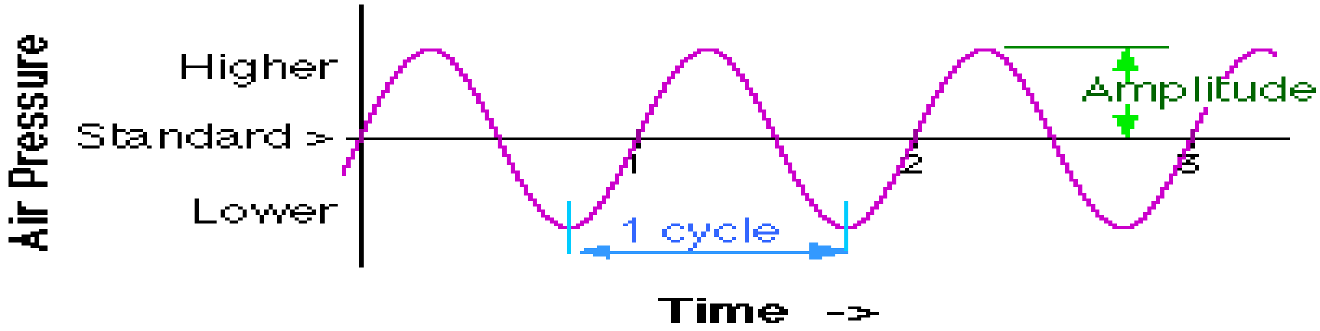
Agenda

- A video is a sequence of images
- A sound is characterised by its frequency (pitch) and amplitude (loudness)
- CD standard quality is 44,100 Hz (sampling) and 16 bits (quantisation)
- Speech signals contain 3 types of sound, some of them are used for speech recognition
- MIDI format for music stores information such as instrument specification, beginning and end of a note, basic frequency, etc.

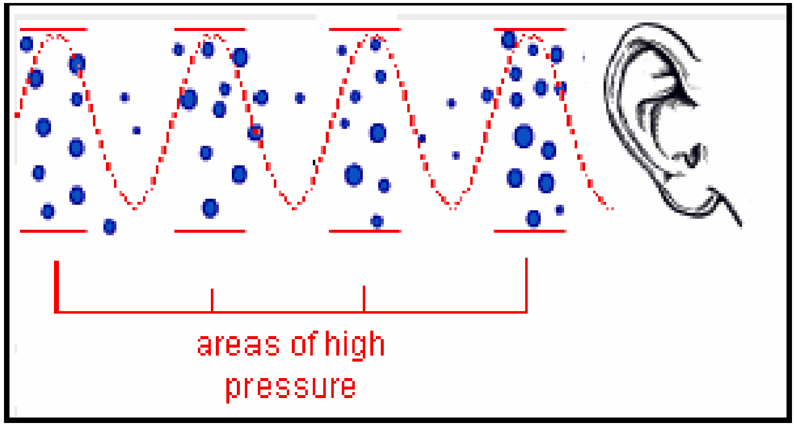
Sound

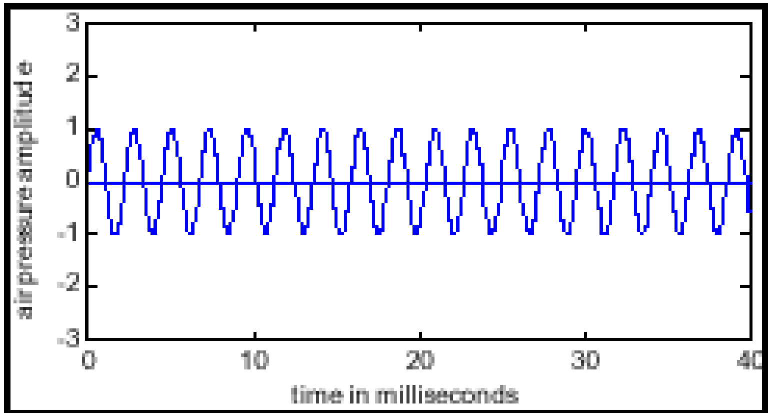
- Sound is a physical phenomenon produced by the vibration of matter, such as a violin string, or a block of wood.
- As the matter vibrates, pressure variations are created in the air surrounding it.
- This alteration of high and low pressure is propagated through the air in a wave-like motion.



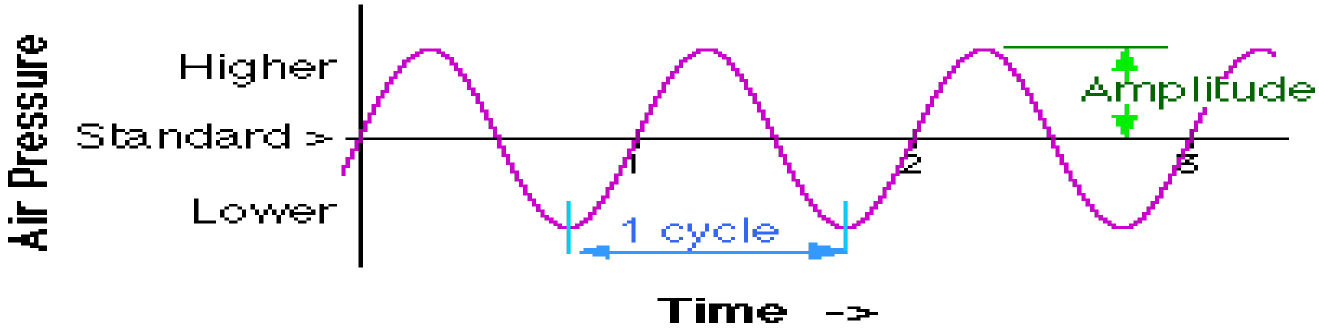


Sound in the analogue domain





Characteristics of Sound Waveforms



- Frequency determines the pitch
  (higher frequency = higher pitch)
  - Infra-sound: from 0 to 20 Hz
  - Human hearing frequency range: 20 Hz – 20 kHz
  - Ultrasound: from 20 kHz to 1 GHz
- Amplitude of the wave determines the volume or intensity (a property subjectively heard as loudness).

Agenda

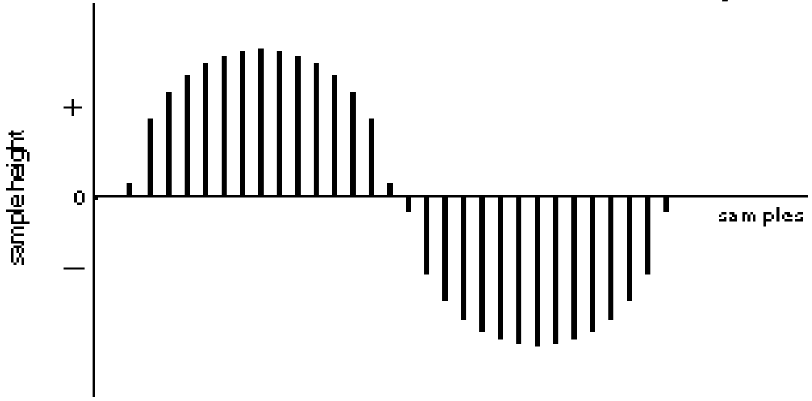
- A video is a sequence of images
- A sound is characterised by its frequency (pitch) and amplitude (loudness)
- CD standard quality is 44,100 Hz (sampling) and 16 bits (quantisation)
- Speech signals contain 3 types of sound, some of them are used for speech recognition
- MIDI format for music stores information such as instrument specification, beginning and end of a note, basic frequency, etc.

Computer Representation of Sound - Sampling -

- A computer measures the amplitude of the waveform at regular time intervals to produce a series of number (sampling). This is done by an ADC (*Analog-to-Digital Converter*)

- Sampling rate: the rate at which a waveform is sampled.

  e.g. *the CD standard sampling rate of 44100 Hz means that the waveform is sampled 44100 times / second*.

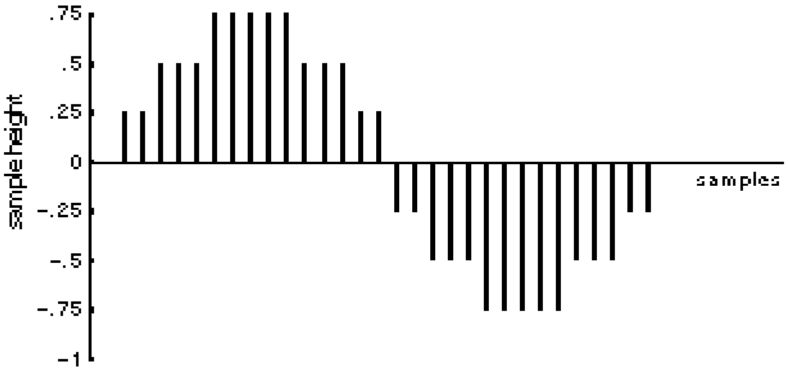


Sampled waveform

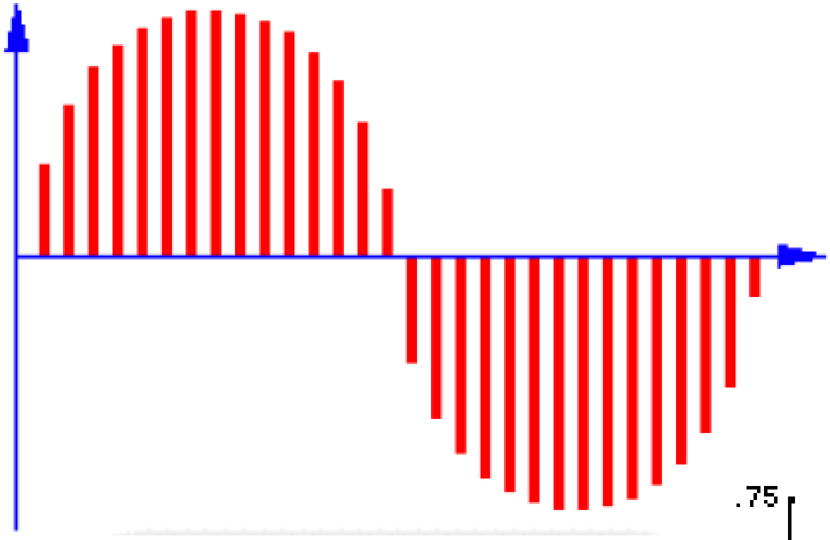
Computer Representation of Sound - Quantisation -

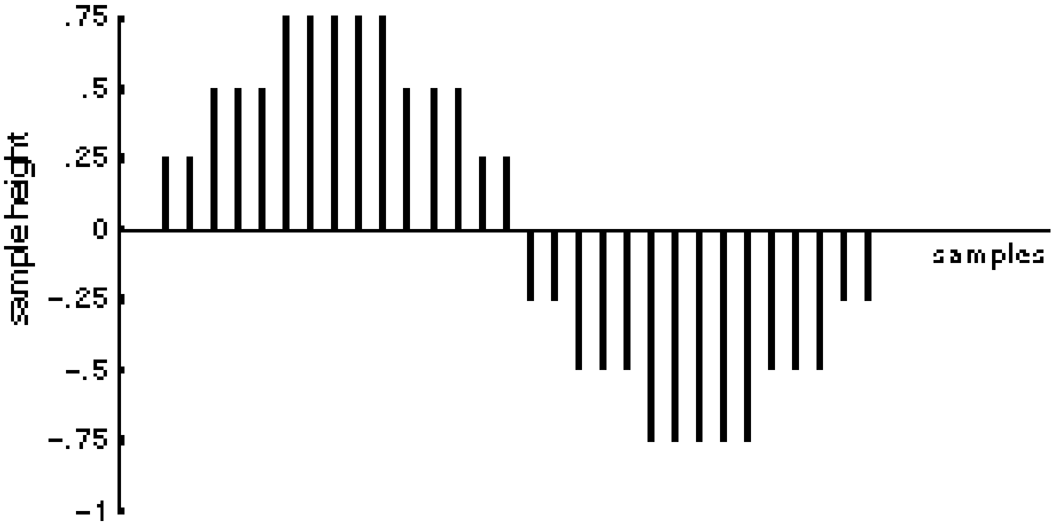
- Quantisation: the resolution or quantisation of a sample value depends on the number of bits used in measuring the height of the waveform (*usually 8-bit or 16-bit*)

3-bit quantisation



Digitisation of Sound





Exercise

A high-quality (CD standard at 44.1KHz) audio signal with 2 channels of 16-bit samples is transmitted uncompressed over an ISDN 64Kbps communication channel.

i) Calculate the number of seconds taken to transmit a one-second burst of audio

ii) Estimate what compression ratio would be needed to transmit the audio in real-time.

Reminder: Nyquist theorem

Sample twice as often as the highest frequency you want to capture

Let f be the frequency of a sine wave. Let r be the minimum sampling rate that can be used in the digitisation process such that the resulting digitised wave is not aliased. Then:

$$r = 2f$$

r is called the ***Nyquist rate***.

Nyquist Rate and Nyquist Frequency

- Given an actual frequency to be sampled, the ***Nyquist rate*** is the lowest sampling rate that will permit accurate reconstruction of an analog digital signal.
- Given a sampling rate, the ***Nyquist frequency*** is the highest actual frequency component that can be sampled at the given rate without aliasing.
- Based on the Nyquist theorem, the Nyquist frequency is half the given sampling rate.

Nyquist Rate and Nyquist Frequency

KEY EQUATION

Given $f_{\max}$, the frequency of the highest-frequency component in an audio signal to be sampled, then the ***Nyquist rate***, f_{nr}, is defined as

$$f_{nr} = 2f_{\max}$$

KEY EQUATION

Given a sampling frequency f_{samp} to be used to sample an audio signal, then the ***Nyquist frequency***, f_{nf}, is defined as

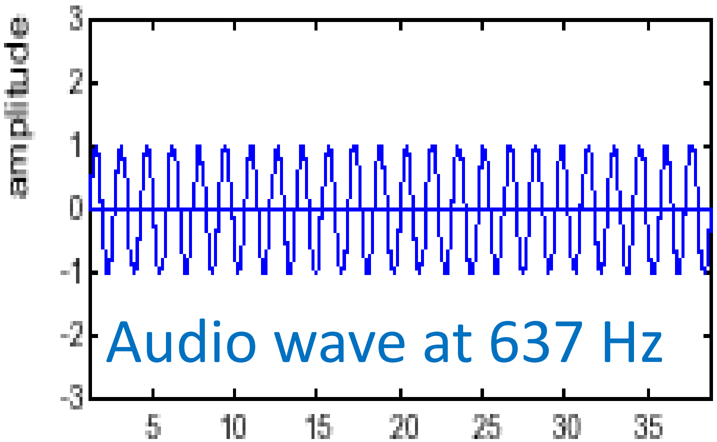
$$f_{nf} = \frac{1}{2}f_{samp}$$

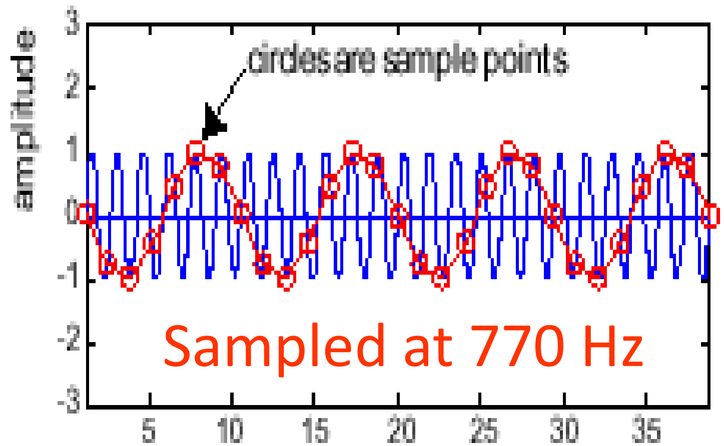
Exercise

The bandwidth of a music signal is between 15 Hz and 20 KHz, assuming the Nyquist sampling rate is used, with 16 bits per sample:

- Derive the bit rate that is generated by the digitisation procedure
- What is the memory in Mbytes required to store a 10 minute passage of stereophonic music?

Aliasing (sampling error)





- The reason a too-low sampling rate results in aliasing is that there aren't enough sample points from which to accurately interpolate the sinusoidal form of the original wave.
- If we take ***more*** than two samples per cycle on an analog wave, the wave can be precisely reconstructed from the samples.

Measuring Sound Amplitude in Decibels

- A decibel is not an absolute unit of measurement.
- A decibel is always based upon some agreed-upon reference point, and the reference point varies according to the phenomenon being measured.
- For sound, the reference point is the *air pressure amplitude for the threshold of hearing*.
- A decibel in the context of sound pressure level is called ***decibels-sound-pressure-level*** (***dB_SPL***).

Measuring Sound Amplitude in Decibels

KEY EQUATION

Let E be the pressure amplitude of the sound being measured and E_0 be the sound pressure level of the threshold of hearing. Then ***decibels-sound-pressure-level***, (***dB_SPL***) is defined as

$$dB\_SPL = 20 \log_{10}\left(\frac{E}{E_0}\right)$$

$E_0 = 0.00002$ Pa

Exercise

- What would be the amplitude (in decibels) of the audio threshold of pain, given as 30 Pa?
- What would be the pressure amplitude of normal conversation, given as 60 dB?

Measuring Sound Amplitude in Decibels

- dB_SPL is an appropriate unit for measuring sound because the values increase logarithmically rather than linearly.
- This is a better match for the way humans perceive sound.
- Experimentally, it has been determined that if you increase the amplitude of an audio recording by 10 dB, it will sound about twice as loud.
- For most humans, a 3 dB change in amplitude is the smallest perceptible change.

Measuring Sound Amplitude in Decibels

Approximate decibel levels of common sounds:

Sound	Decibels (dB_SPL)
Threshold of hearing	0
Rustling leaves	20
Conversation	60–70
Jackhammer	100 (or more)
Threshold of pain	130
Damage to eardrum	160

Signal to Quantisation Noise Ratio (SQNR)

- SQNR is also measured in decibels.
- SQNR is directly related to ***dynamic range:*** the ratio of the largest sound amplitude and the smallest that can be represented with a given bit depth.

Let *n* be the bit depth of a digitised media file (e.g. digital audio). Then the signal-to-quantisation noise ratio ***SQNR (or dynamic range)*** is:

$$SQNR = 20 \log_{10}(2^n) = 20n \log_{10}(2) \sim= 6n$$

Dynamic Range

KEY EQUATION

Let n be the bit depth of a digital audio file. Then the ***dynamic range of the audio file***, d, in decibels, is defined as

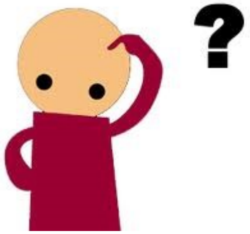
$$d = 20n \log_{10}(2) \approx 6n$$

- You can estimate that an *n*-bit digital audio file has a dynamic range (or, equivalently, a signal-to-noise-ratio) of 6*n* dB.
- Dynamic range is a relative measurement—the relative difference between the loudest and softest parts representable in a digital audio file, as a function of the bit depth.

Exercise

- What is the dynamic range (SQNR) of a 16 bit digital audio file?
- How about a 8 bit digital audio file?

Question



A sound file encoded with a 8 bits quantisation rate is likely to be:

- A piece of music
- Natural sound (e.g. rain)
- Speech
- A song

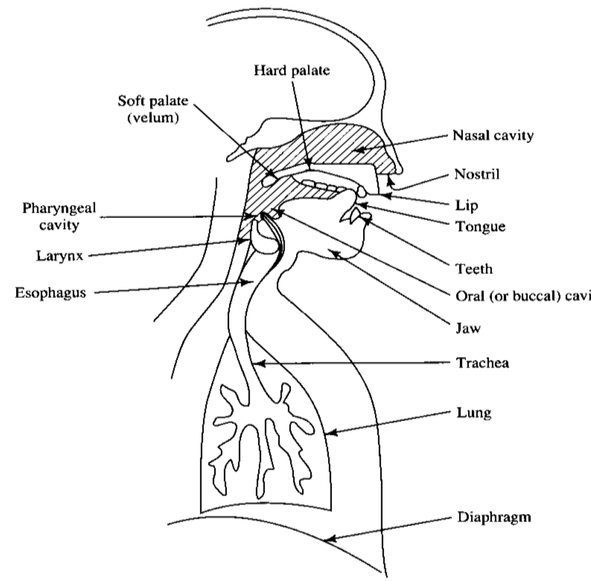
Quantisation Error

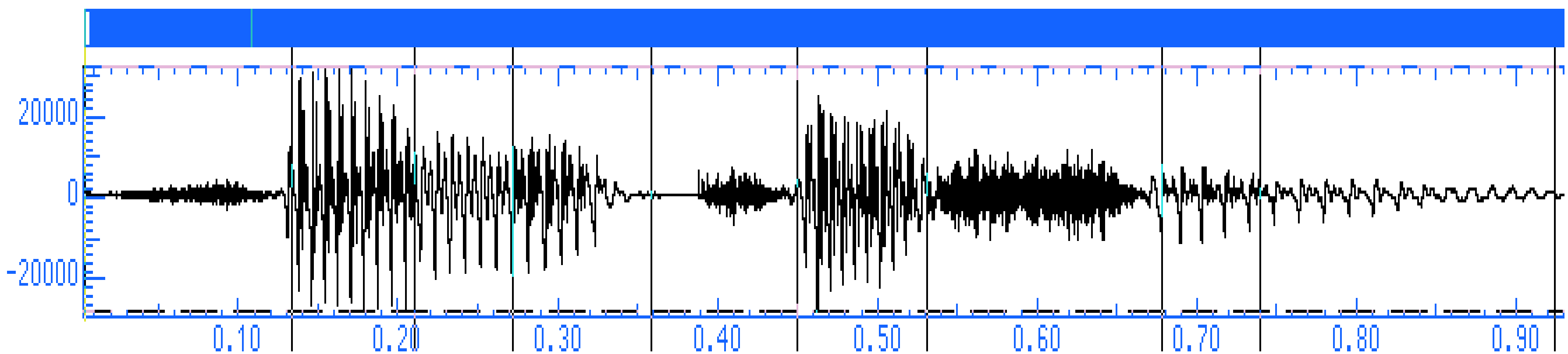
- While an insufficient sampling rate can lead to aliasing, an insufficient bit depth can create quantisation error.
- ***Audio dithering*** is a way to compensate for quantisation error. The way to do this is to add small random values to samples in order to mask quantisation error.
- ***Noise shaping*** is another way to compensate for the quantisation error: it redistributes the quantisation error so that the noise is concentrated in the higher frequencies, where human hearing is less sensitive

Agenda

- A video is a sequence of images
- A sound is characterised by its frequency (pitch) and amplitude (loudness)
- CD standard quality is 44,100 Hz (sampling) and 16 bits (quantisation)
- Speech signals contain 3 types of sound, some of them are used for speech recognition
- MIDI format for music stores information such as instrument specification, beginning and end of a note, basic frequency, etc.

Speech

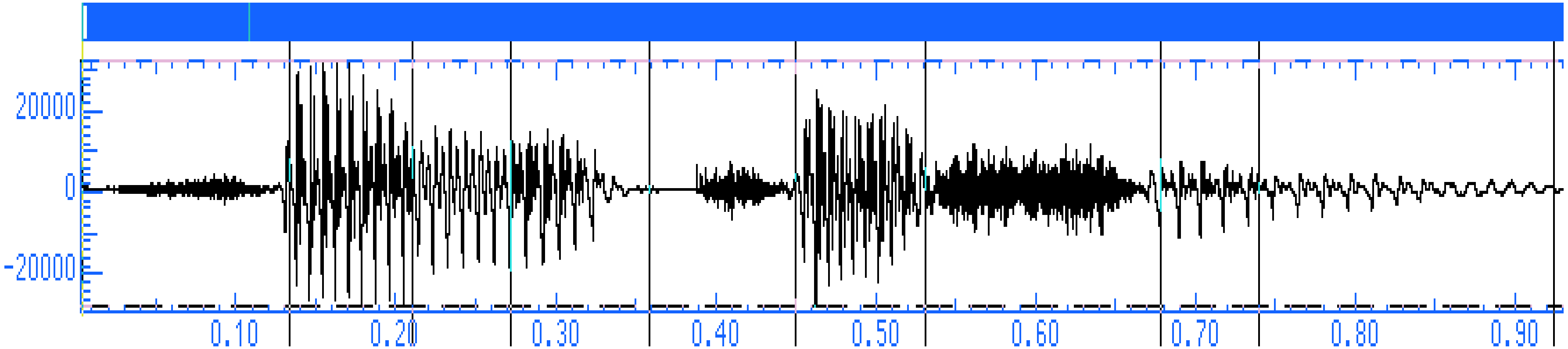




f o n e t i c ia n

Types of Speech Sounds

- Voiced sounds : the vocal chords are vibrated, which can be felt in the throat. All vowels are voiced.
- Fricatives (unvoiced sounds) : a consonant, such as *f* or *s* in English, produced by the forcing of air through a constricted passage.
- Plosives (also unvoiced sounds) : a speech sound produced by complete closure of the oral passage and subsequent release accompanied by a burst of air, as in the sound (d) in *dog.*



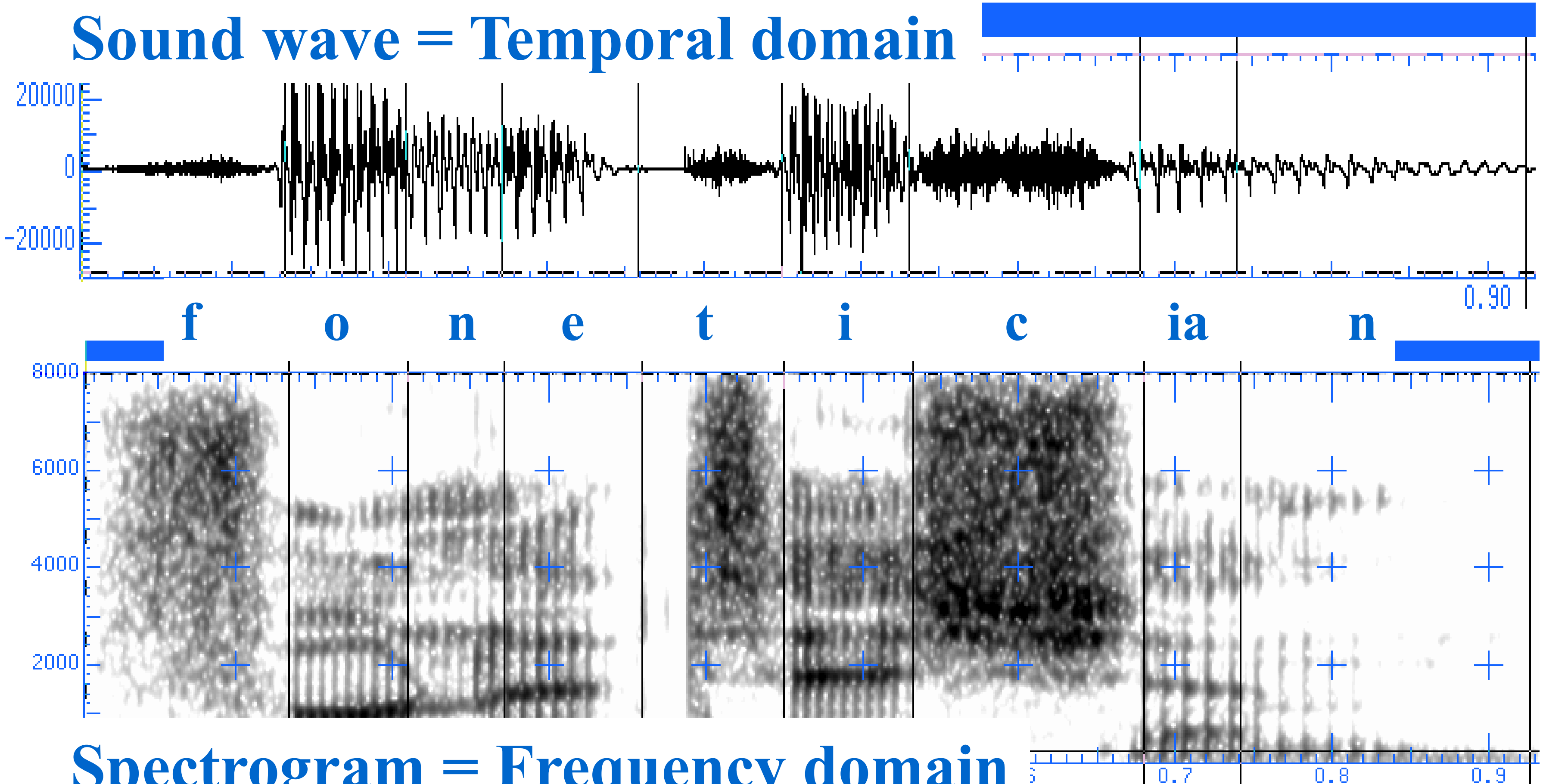
Voiced Speech Sounds

Voiced speech sounds have two properties which can be used in speech processing:

- Speech signals show during certain time intervals almost periodic behaviours. These signals are *quasi-stationary signals* for around 30 ms.
- The spectrum of speech signals (voiced sounds) shows characteristic maxima. These maxima, called formants, occur because of resonances of the vocal tract.

Temporal and Frequency Domains

Sound wave = Temporal domain



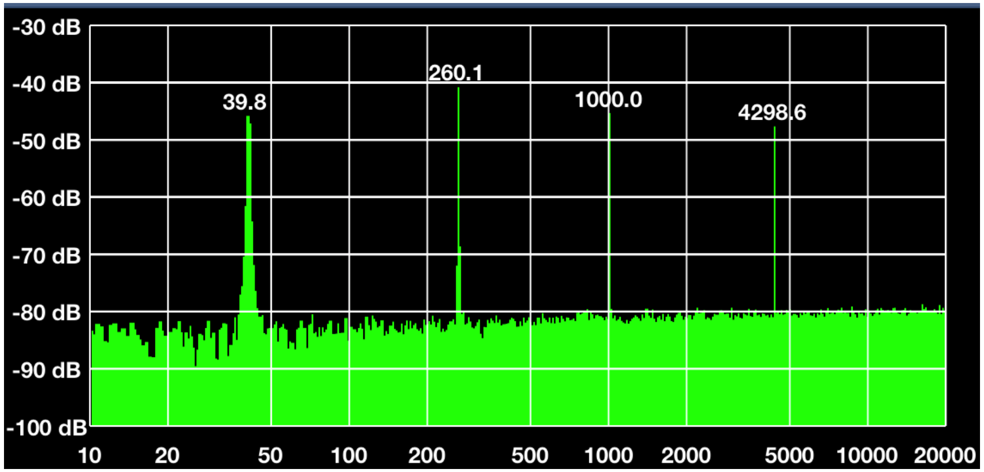
Spectrogram = Frequency domain

Temporal and Frequency Domains

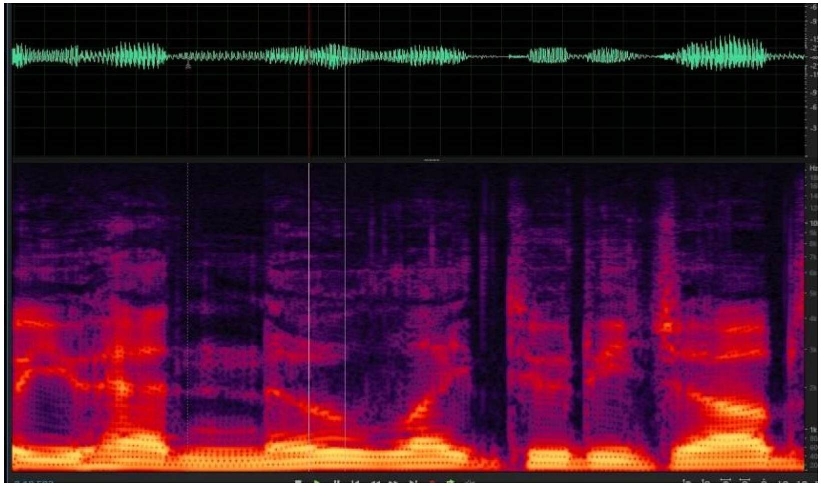
- Sound can be represented either over the time domain or the frequency domain.
- The transform that is used most frequently in digital audio processing is the Fourier transform.
- A complex waveform is equal to an infinite sum of simple sinusoidal waves, beginning with a ***fundamental frequency*** and going through frequencies that are integer multiples of the fundamental frequency.
- These integer multiples are called ***harmonic frequencies***.
- In the ***frequency domain***, data is stored as the amplitudes of frequency components

Frequency Domain

A Power Spectrum is a 2-dimensional representation (frequency(x) / amplitude(y); a Spectrogram is a 3-dimensional representation (time(x) / frequency(y) / amplitude(colour).



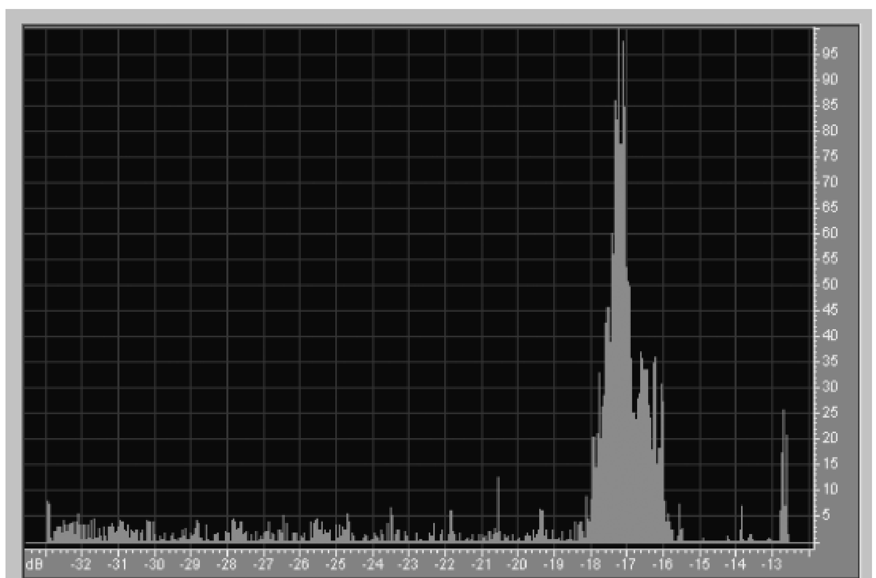
Power Spectrum



Spectrogram

Audio Histogram

- Audio processing programs sometimes offer a statistical analysis of your audio files, which analyse sample values in the **time domain**.
- An ***audio histogram*** shows how many samples there are at each amplitude level in the audio selection.



Speech Processing Applications

Speech Analysis

- Who?
  - Verification
  - Identification
- What?
  - Recognition
  - Understanding
- How?

Agenda

- A video is a sequence of images
- A sound is characterised by its frequency (pitch) and amplitude (loudness)
- CD standard quality is 44,100 Hz (sampling) and 16 bits (quantisation)
- Speech signals contain 3 types of sound, some of them are used for speech recognition
- MIDI format for music stores information such as instrument specification, beginning and end of a note, basic frequency, etc.

Music

• Thus far, we've been considering digital audio that is created from sampling analog sound waves and quantising the sample values.

•There's another way to store sound in digital form: ***MIDI (Musical Instrument Digital Interface)***.

• MIDI stores "sound events" or "human performances of sound" rather than sound itself.

• The difference between sampled digital audio and MIDI is analogous to the difference between bitmapped graphics and vector graphics

MIDI Data Format

- A MIDI file contains messages indicating when a note begins (Note On), when a note ends (Note Off), what the note is, how hard it is pressed (Velocity), how hard it is held down (Aftertouch), what instrument is played, and so forth.
- Each MIDI message communicate one musical event.

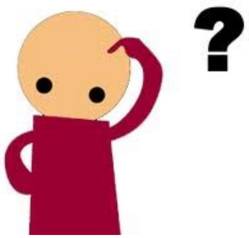
e.g. when a musician presses a piano key, the MIDI interface creates a MIDI message where the beginning of the note with its stroke intensity is encoded.

- The MIDI standard identifies 128 instruments (including noise effects) with unique numbers (e.g. 41 for the violin).

MIDI Hardware and Software

- Hardware devices that generate MIDI messages are called ***MIDI controllers***.
- Devices that read MIDI messages and turn them into audio signals are called ***MIDI synthesizers***. Two methods for synthesizing sound are ***frequency modulation synthesis*** and ***wavetable synthesis***.
- A ***MIDI sequencer*** is a hardware device or software application program that allows you to receive, store, and edit MIDI data.

Questions



- Why are MIDI encoded music signals very small?
- What other advantage to MIDI audio is there compared to sampled digital audio?
- Is there any disadvantage to MIDI audio compared to sampled digital audio?

Musical Acoustics and Notation

- In Western music notation, musical sounds, called ***tones,*** are characterized by their pitch, timbre, and loudness.
- With the addition of onset and duration, a musical sound is called a ***note***.
- The ***pitch*** of a note is how high or low it sounds to the human ear.
- The ***timbre*** of a musical sound is its “tone color”.
- The lowest frequency of a given sound produced by a particular instrument is its **fundamental frequency**. Then there are other frequencies combined in the sound, which are integer multiples of the fundamental frequency, referred to as ***harmonics.***

Exercise

- A musical note played on an instrument consists of a fundamental frequency and, depending on the instrument, different numbers of harmonics. Each harmonic is an integer multiple of the fundamental frequency.
- Given a note from a musical instrument, which contains only the following frequency components: 100Hz, 200Hz, 300Hz, and 400Hz, **at what rate would you need to sample this sound to ensure that the sampled audio was of the same fidelity as the original note?**
- Assuming that the amplitude of each harmonic is half the amplitude of the previous harmonic, **sketch the signal in the frequency domain for the above note.**

Musical Acoustics and Notation

If the frequency of one note is 2^n times of the frequency of another, where n is an integer, the two notes sound "the same" to the human ear, except that the first is higher-pitched than the second.

KEY EQUATION

Let g be the frequency of a musical note. Let h be the frequency of a musical tone n ***octaves*** higher than g. Then

$$h = 2^n g$$

Exercise

If the frequency of a note A is about 440 Hz, what is the frequency of an A two octaves below the 440 Hz A?

Summary

- A video is a sequence of images
- A sound is characterised by its frequency (pitch) and amplitude (loudness)
- CD standard quality is 44,100 Hz (sampling) and 16 bits (quantisation)
- Speech signals contain 3 types of sound, some of them are used for speech recognition
- MIDI format for music stores information such as instrument specification, beginning and end of a note, basic frequency, etc.