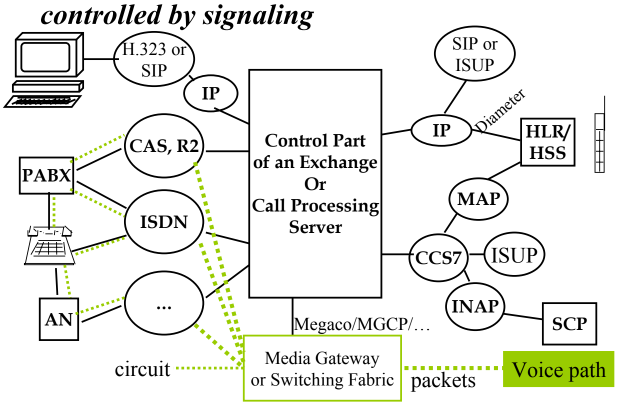
Voice Coding, PCM Voice, Voice Quality, E-model

- PCM ~ Pulse Code Modulation
  - Sampling
  - Quantizing
    - Linear
    - Non-linear
  - Quantizing error
- PCM frame structure
- Other Voice coding algorithms
- E-model, Voice quality measurements
- Requirements to signaling.

Voice path is established and its quality is controlled by signaling



Requirements for the Voice path and the Switching Fabric

- In CSN the Switching Fabric must understand the bits, the timeslots and the frames in the same way as the transmission systems that carry the bits
  - The Fabric and the transmission systems must be synchronized
- Voice must be coded efficiently (what is efficient changes over time)
- An exchange must supervise voice connections:
  - calls shall/should not be offered to faulty connections
  - calls must sometimes be cleared from faulty connections
  - detected faulty connections must be reported to the far end if possible
- In a packet network voice path supervision is delegated to terminals
  - Many routers are unable to detect link failures with hardware. Instead the routing protocol hello messages are used → slow error detection and packet loss.
  - Signaling still must control the creation of the path e.g. for traversing NATs!

Sampling

- Nyquist theorem
  - If an analogue signal with limited spectrum is sampled regularly with a frequency of at least twice as high as the highest frequency component, the samples carry all the information in the original signal. The original signal can be reconstructed using a low pass filter.

Key assumptions in Circuit telephony: PSTN, ISDN

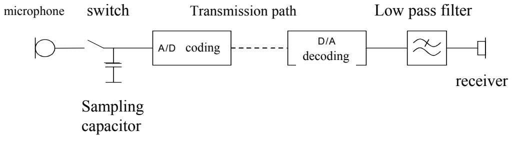
- In voice transmission, the spectrum carried is specified to be 300 - 3400 Hz, resulting in a minimum sampling rate of 6,8 kHz.
- In practice, since the width of the transmission channel in an analogue system is 4kHz, in a digital system a sampling rate of 8 kHz (8000 samples/s) is used.

Wideband codecs, such as WB-AMR and WB-GIPS use a sampling rate of 16 kHz

WB –wideband
GIPS Global IP Sound –used in Skype

Digital voice transmission

- The voice path includes a microphone, A/D-converter, D/A-converter and a loudspeaker.
- In practice, the analogue signal needs to be filtered before the conversion

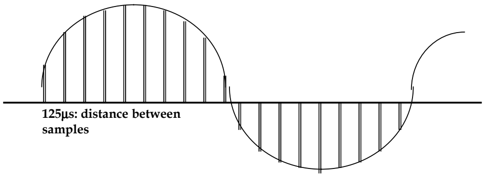


Pulse Code Modulation - PCM

- In PCM, analogue voice is digitized and thus it can be carried by digital transmission systems and switched in digital switching fabrics.
- PCM was invented in 1937 but the first real implementations became possible only with transistor technology during 1960's. This is also one of the origins of Nokia Electronics (1968) and Nokia Networks.
- PCM conversion has four steps:
  - filtering
  - sampling
  - quantizing
  - coding

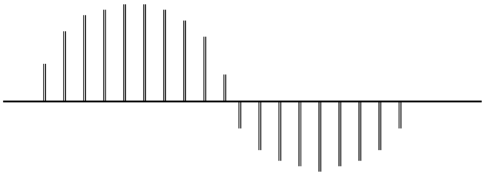
Sampling of the analogue signal

- Sampling of the analogue signal is done with a frequency of 8 kHz, I.e. with an inter-sample interval of 125 µs.
- The result is a PAM –signal: 8000 samples/second evenly spaced in time



Pulse Amplitude Modulation PAM

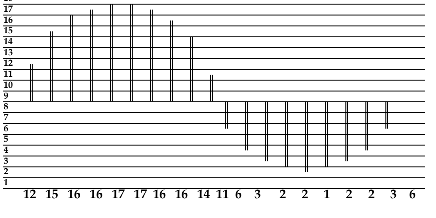
- Sampling produces a time discrete PAM signal reflecting the amplitude of the analogue signal.
- PAM-signal is quantized producing PCM-code.



Quantizing = replacement of real value by the closest integer.

Quantizing results in approximation of the samples

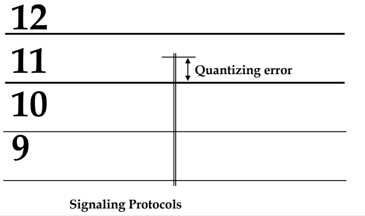
- Real valued amplitude figures are replaced by discrete integer values.
- Quantizing should result in values that appear in the signal with equal probability.



Quantizing distortion

- Quantizing produces distortion, that is called quantizing distortion.
- Quantizing distortion is made by the replacement of real values by their integer approximates and at maximum can reach ½ quantizing interval.
- In linear quantizing the signal to distortion ratio is

S/D=6n+1,8 dB n=word length



Linear vs. non-linear

- The result of quantizing should use signal values with equal probability.
- This results in minimization of distortion, because a larger number of discrete signal values falls into the most typical analogue signal value area.
- The effect of the quantizing error on voice quality is averaged over time by the human ear.
- In a voice signal, small analogue values appear with higher probability than larger values.

--> non-linear quantizing

Non-linearity

- Non-linear conversion can be implemented in two ways:
  - using non-linear quantizing
  - using compression before linear quatizing is applied
- Non-linear quantizing can be implemented e.g. using a network of resistors, compression requires a non-linear amplifier.
- Irrespective of the method of implementation, the non-linear quantizing follows a conversion function giving the mapping of analogue signal values to integers.
  - In Europe (ETSI) A-function
  - In USA (ANSI) µ-function

PCM-coding and quantizing

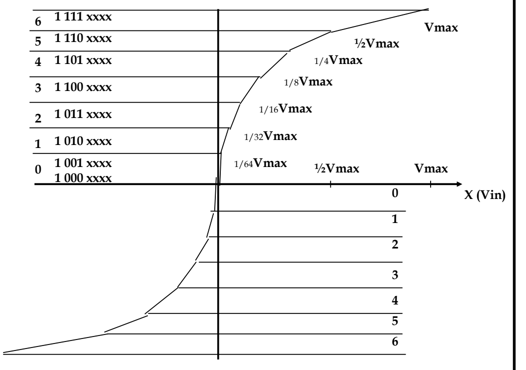
- ✓ Accoding to ETSI specification, voice coding uses 8 bits per sample.
  - bit-1 gives the polarity of the signal
  - bits 2-4 give the segment of the non-linear quantizing
  - bits 5-8 give the value of the discrete signal inside the segment
- ✓ Non-linearity follows the so called A -law

$$\left(\frac{A|x|}{1+\ln(A)}\right), 0 \le |x| \le \frac{1}{A}$$

The value of A is 87,6.

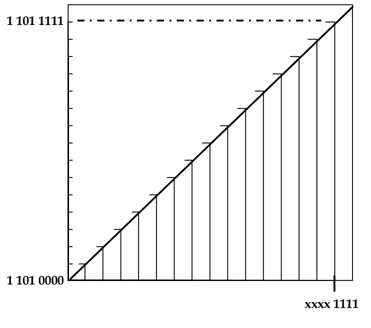
$$\left(\frac{1+\ln|Ax|}{1+\ln(A)}\right), \frac{1}{A} \le |x| \le 1$$

Quantizing according to the A-law



Quantizing inside a Segment

- ✓ In a segment quantizing is linear



Linear vs non-linear quantizing

- ✓ Linear and non-linear quantizing can be compared using the gain in signal resolution by non-linearity.
- ✓ Non-linear quantizing emphasizes small signal values, for which a gain in resolution of 24 dB is achieved.

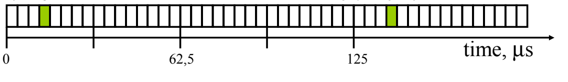
$$\boxed{G_{dB}=20\log V_{in}/V_{comp}}$$

PCM-hierarchy

- ✓ PCM-hierarchy is created by interleaving time division multiplexed signal connections byte by byte (sample by sample). Bits become shorter.
- ✓ The basic speed in the hierarchy is the bitrate of a single voice channel

S=8000Hz* 8bit = 64kbit/s,

- ✓ in time in a 2Mbit/s PCM system, this looks like:



- ✓ The following higher order PCM systems are defined
  - 30 voice channels
  - 120 voice channels
  - 480 voice channels
  - 1920 voice channels

Higher order PCM systems still provide 64kbit/s voice channels for each call, but there are more of them on a connection. Bits and octets become shorter in time.

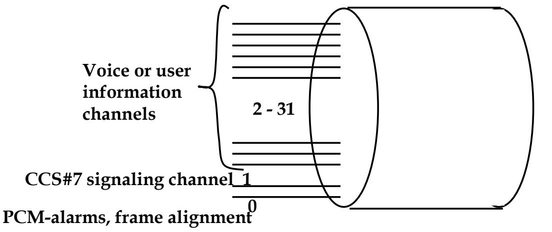
PCM 30 (E1)

- ✓ The most common information switching and transmission format in the telecommunication network is PCM 30.
- ✓ PCM 30 contains:
  - 1 synchronization and management channel
  - 1 signaling channel
  - 30 voice channel
- ✓ A channel is a time slot in the PCM-frame (125µs), created by TD multiplexing.
- ✓ PCM 30 system carries 32 time slots, each 64kbit/s. This gives a total bit rate of 2048kbit/s.

PCM 30 frame

- ✓ **PCM 30 -frame contains 32 time slots**
  - › **time slot 0 is dedicated for synchronization and management information**
  - › **Time slot 16 is assigned for signaling information (CAS). In ISDN also TLS 16 can be used for e.g. voice transfer.**
  - › **Time slots 1-15 and 17-32 are voice or user information channels, this means that e.g. voice bit of 30 simultaneous voice conversations can be multiplexed onto a single PCM30 connection. Each voice channel uses the speed of 64 kbit/s.**
- ✓ **Even and odd frame structures differ**
  - › **In even numbered frames, time slot 0 carries the frame alignment signal (C0 01 10 11). C is the CRC-bit (cyclic redundancy check) for ensuring the frame alignment recovery in case someone is sending X0 01 10 11 on a user information channel – this addition was forced by ISDN which supports transparent 64kbit/s service for data transfer.**
  - › **Time slot 0 in odd frames carries alarm information. To avoid wrong frame alignment, the second bit in tsl 0 is set to the constant value of 1.**

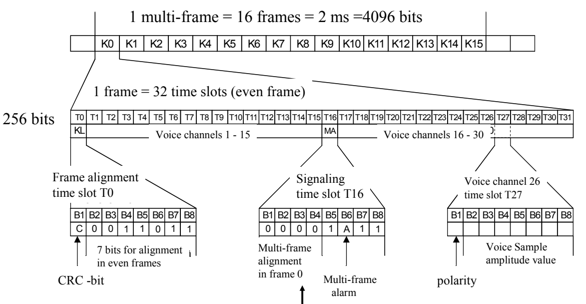
The use of PCM time slots in the Finnish CCS#7 network



Nowadays, tsl 16 is used for voice!
On PCM:s that do not need to have a signaling channel, Tsl-1 may be used for voice or left reserved for signaling for simplicity.

Even numbered PCM 30 -frame

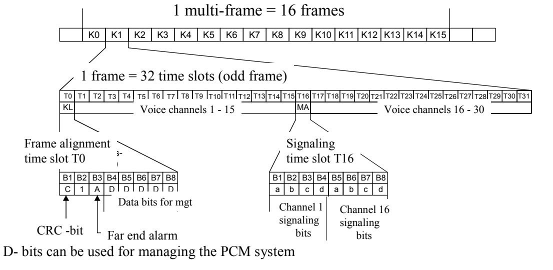
Speed is 2,048 Mbit/s



C bit carries a checksum calculated over previous several frames

Applies only to K0, other even numbered, look at the next slide

PCM-frame structure (odd frame)



Actually, this system has two independent multi-frame structures. One based on TSL 16, the other is based on the CRC bit: the first half of the multiframe (2048 bits) is used to calculate the 8-bit CRC and the second half of the multiframe is used to carry the CRC. Nowadays, most times the TSL 16 multiframe structure is historical only.

The CRC story in PCM

- ✓ **PCM frame synchronisation is based on the idea that TSL 0 contains a certain bit combination (octet value). When a PCM system is started, the receiver sees a bit stream and does not know timeslot or frame boundaries.When the receiver sees the synch value (x0011011) several times with the known distance from each other, it decides that it has found TSL 0 and is able to number the rest of the TSLs simply by counting octets. The receiver recognises odd and even frames based on the B2 bit value in TSL 0 of an odd frame.**
- ✓ **In the analogue era, only voice was sent in voice TSLs, so the particular octet value for frame synchronisation could not appear consistently in a particular voice time slot (very low probability). One such appearance is not a problem.**
- ✓ **When ISDN was introduced with unrestricted 64kbit/s bearer service (=transparent 64kbps channel for a user), it became possible even for the fun of it to send the frame synch octet value in a "voice channel" consistently. If now the PCM system goes to initial synchonization mode, it could lock into a user channel and mess up the TSLs between users. This was not acceptable.**
- ✓ **So a fix was invented. CRC bit was added into TSL 0. It contains a checksum over all timeslots of 8 consecutive PCM frames and is sent in 8 consecutive frames in the CRC bit. So, 8+8 frames form the multiframe structure in the PCM systems used in ISDN networks. This CRC function is implemented on silicon in Exchange terminal in Digital Exchanges. Now a receiver is able to make a difference between x0011011 in TSL0 and the same 7 bits in a user channel based on the CRC (bit nr 1 in TSL0). To fool the system, a malicious user would have to see the contents of other timeslots to also calculate the CRC bit correctly.**

A number of other voice coding algorithms exist, more are developed all the time.

- ✓ **PCM coding is called G.711 – an ITU-T standard**
- ✓ **Examples:**
  - › **GSM EFR codec (enhanced full rate),**
  - › **AMR (Adaptive Multirate) is the new emerging cellular standard codec, has NB-AMR and WB-AMR variants – narrow band, wide band. Wide band means that Voice is first cut into 7kHz (not 4kHz) prior to sampling.**
  - › **G.7xx – many codecs for packet voice, many of them patented, patents require licensing – difficult to use widely.**
- ✓ **Leads to a need to negotiate about codecs end-to-end! This is a requirement for signaling. In CS networks, a codec needs to be standardized globally. In PS networks, it is enough to agree on a small set and be able to agree on a common codec end to end for a call.**

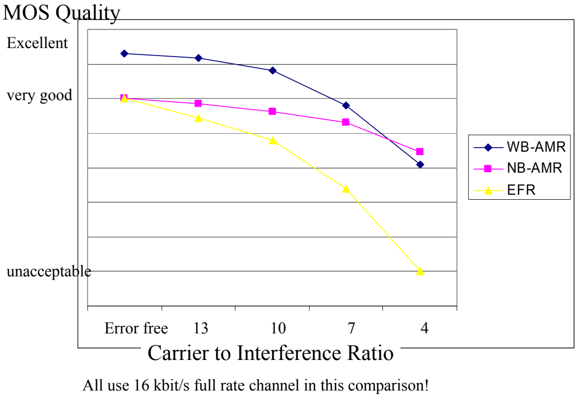
Some codecs and their characteristics

Coding Standard	Algorithm	Sample Size (msec)	Rate Kbit/s	Mean Opinion Score	Year
G.711	PCM	0.125	64	4.10	1972
GSM 06.10	RPE-LTP	20.000	13	3.50	1987
G.726,G727	ADPCM	0.125	16, 24, 32, 40	3.85	1990
G.728	LDCELP	0.625	16	3.61	1992, 1994
IS-96	VSELP	20.000	8.5, 4, 2, 0.8		1993
G.729, G.729a	CS-ACELP	10.000	8	3.92, 3.70	1995
G.723.1	MPC-MLQ	10.200	6.3, 5.3	3.90	1995
PDC	PSI-CELP	40.000	3.45		1996
FS-1015	LPC	25.700		2.40	-
AMR-NB					
AMR-WB				>PCM	

Voice quality can be assessed by Mean Opinion Score or MOS -value

- ✓ **Take 20 people, organise a controlled experiment with recorded voice samples (both male and female voices), use several languages,**
  - › **After listening the test subject marks his/her opinion: 5 –excellent quality, 1 – bad quality, Repeat for many samples, Calculate averages.**
  - › **Make sure people do not get bored, so same people can not be used for long.**
  - › **Results may depend on time, test conditions and the group of people**
  - › **Method is also called Absolute Category Rating**
- ✓ **Alternatively a comparative method can be used – Poor or Worse (PoW), Good or Better (GoB)**
- ✓ **Cumbersome and expensive → objective measurements.**

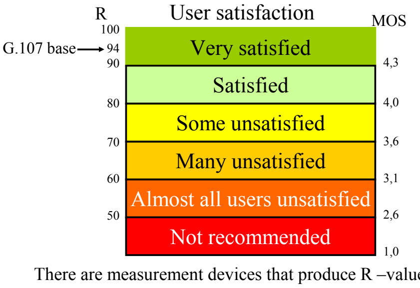
Comparison of GSM and AMR codecs



All use 16 kbit/s full rate channel in this comparison!

E-model (G.107) produces the R-value for characterizing voice quality

- ✓ **R-value varies between 0….100. In practice below 50 is unacceptable quality. With narrow band coding (3.1 kHz band) the maximum R-value is 94.15**



MOS scale is
5 – Excellent
4 – Good
3 – Fair
2 – Poor
1 – Bad

There are measurement devices that produce R –values!

R-value is an ”objective” measure calculated based on voice impairements

Impairements include: packet loss(sample loss), echo, delay, noise, etc

Impairements are additive over a connection!

$$R = R_0 - Is - Id - Ie + A$$

R_0 – basic value reflecting signal to noice ratio
Is – sending impairements
Id – delay and echo impairements
Ie – handware (e.g. codec) impairements (G.113 has a list of values for different codecs
A – reflects positive conditions (mobility, satellite…)

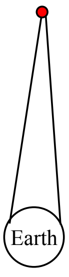
$$MOS = 1 + 0.035R + R(R - 60)(100 - R)*7e\text{-}6$$

Objective vs subjective measurement of voice quality

- ✓ **For an objective method it must be shown that the result reflects well the result of a subjective measurement (correlation analysis over a large number of experiments). So, irrespective of the high price and cumbersomeness of subjective measurements, they tell the truth = perception by real users.**
- ✓ **Objective measurement of R value needs both the original signal and the impaired signal. This means that the method can not be used in the network. It is used in a Lab when systems are developed.**
- ✓ **There are less accurate methods that estimate voice quality based on the impaired signal only e.g. in a packet network. Such methods can be implemented e.g. in a media gateway and used by the operator for network monitoring purposes.**

To eliminate echo on long connections, echo cancellers and echo suppressors are used – these need to be controlled by signaling

Delay example: distance from A to B is 20 000 km in Fiber:



$$\text{Delay}=\frac{20\ 000\ \text{km}}{200\ 000\ \text{km/s}} = 100\ \text{ms}$$

Satellite on the Geostationary orbit:

$$\text{Delay}=\frac{80\ 000\ \text{km}}{300\ 000\ \text{km/s}} = 266\ \text{ms}$$

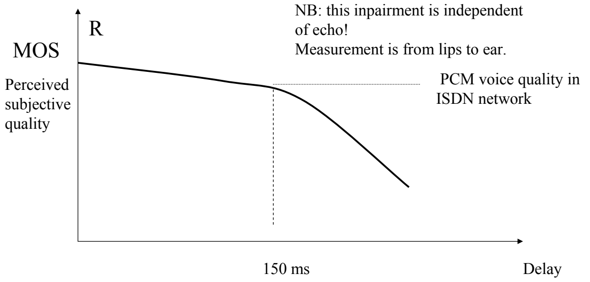
Echo is produced at 4/2 wire conversion. Example is analogue subscriber interface. Also voice can leak from loadspeaker to microphone (speakerphone). When delay > 30 ms, echo needs to be cancelled.

Echo as a voice impairment

- ✓ **Echo = a speaker hears his or her own voice with a delay through the earpiece.**
- ✓ **Can be created in 2/4 wire transformation in case on analogue subscriber lines (subs line = copper pair = 2 wires, while PCM system transfers voice in two directions logically on separate wires). This 2/4 transformation takes place on Analogue line cards in local exchanges in both ends**
- ✓ **If transfer is phone to phone digital, no network echo is created (all the way logically "4 wires") – what remains is acoustic echo = at the remote end voice is acoustically transmitted over the air from the loadspeaker (e.g. by a speaker phone) to the microphone.**
- ✓ **If delay between direct voice and echo > 30 ms, Echo needs to be removed, otherwise significantly degrades call quality. Can be removed either**
  - › **in terminals (e.g. packet networks) or by**
  - › **echo cancellation equipment connected to exchanges on the path in CS networks. In this case commands for turning echo cancellation on/off must come from call control and call control needs to know whether it is dealing with a voice call or data call and whether EC needs to be applied or not for voice calls. This can be found out by having configuration info about "this is a long connection that may need EC) + having info about call type voice/data carried in signaling.**
  - › **EC is implemented with DSPs that scan for a reverse voice pattern that resembles the direct voice bit pattern on the voice channels == is a heavy operation.**

ADD 12.3.2007

Voice quality starts to degrade fast, when one way end-to-end delay > 150ms



Quality can be measured e.g. based on the E-model or using MOS –measurements.
MOS - Mean Opinion Score.

The impact of transcoding

Voice stream in code X → trans-coder → Voice stream in code Y

- ✓ **Conversion from one digital code to another (E.g from GSM EFR to G.711) is called transcoding.**
  - › **implemented usually by decoding to analogue voice and encoding using the other coder**
  - › **direct conversion in digital form is poorly known and there are no general solutions**
  - › **GSM codec to G.711 transcoding takes place in transcoders that are part of BSC but in practice most times reside next to MSC in GSM networks.**
- ✓ **Causes delay and always degrades voice quality.**
- ✓ **Requirement to network signaling: locate the callee in such a way that transcoding is applied only when absolutely necessary and hopefully never twice!**
- ✓ **Signal to establish transcoder free operation when possible.**

Summary: Voice transfer and signaling

- ✓ **Voice path set-up is controlled by signaling**
- ✓ **Voice quality is controlled by switching systems in circuit switched networks**
  - › **e.g. voice path testing prior to call set-up**
  - › **Signaling may carry information that "this is a voice call" and apply echo cancellers on long international connections.**
  - › **Echo cancelling must not be applied to data connections!**
  - › **Coders are globally standardized**
- ✓ **In Packet networks voice quality is an end-to-end matter – terminals are responsible**
  - › **Terminals may also negotiate which coder to use, the network does not need to know about that (end to end signaling)!**
  - › **If terminals do not have a common codec, a transcoder (media gateway) in the network is needed.**
- ✓ **Quality impairements are additive end to end! Better select such path that impairements are minimized.**
  - › **Transcoder Free Operation, Translation Free Operation etc...**