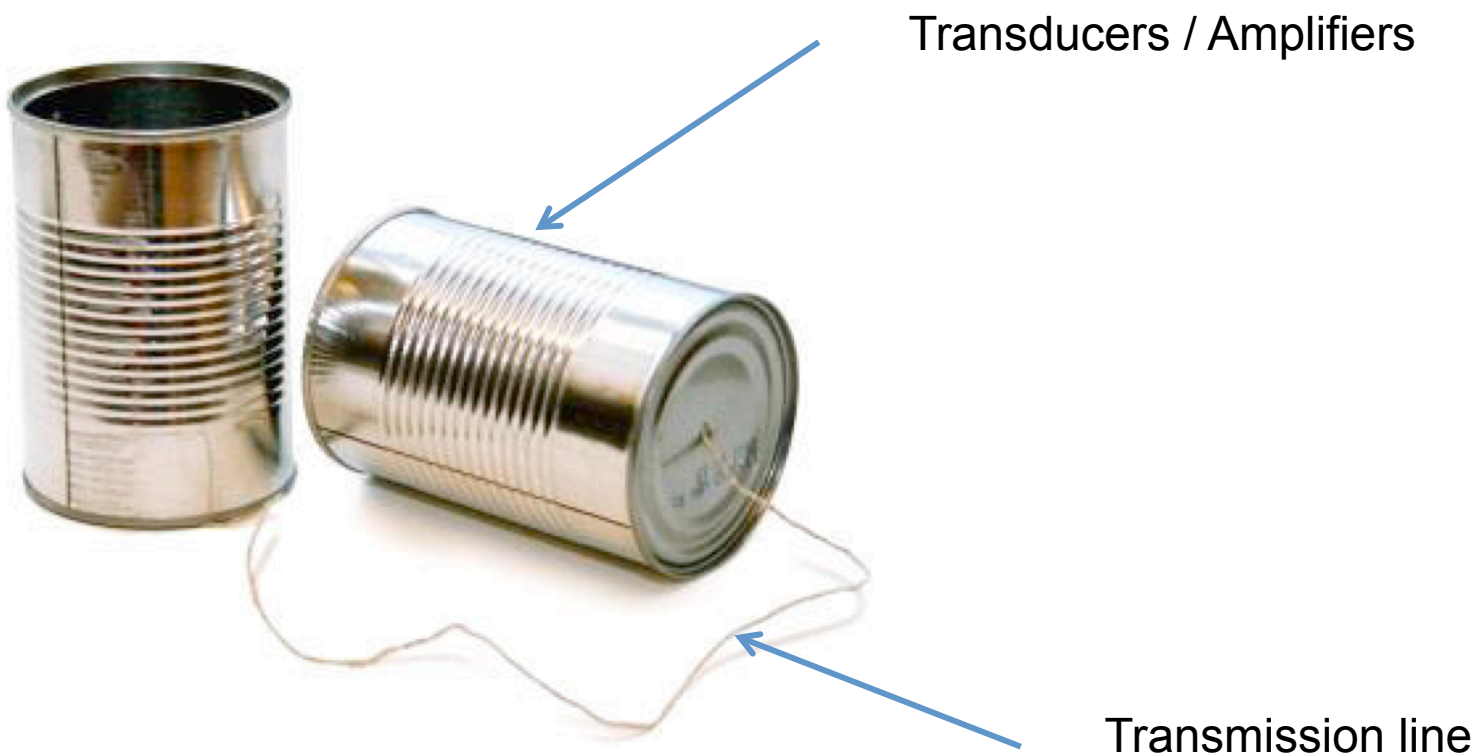
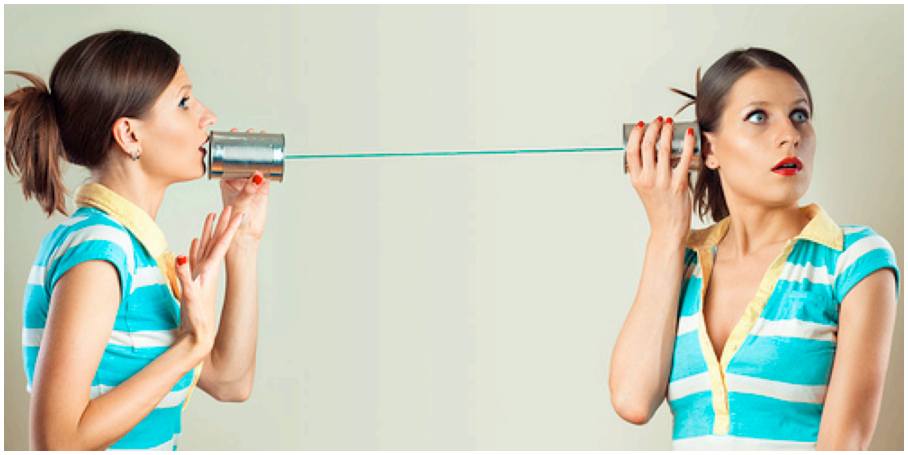
Codec matrix

Michael Knappe
Co-chair, codec WG

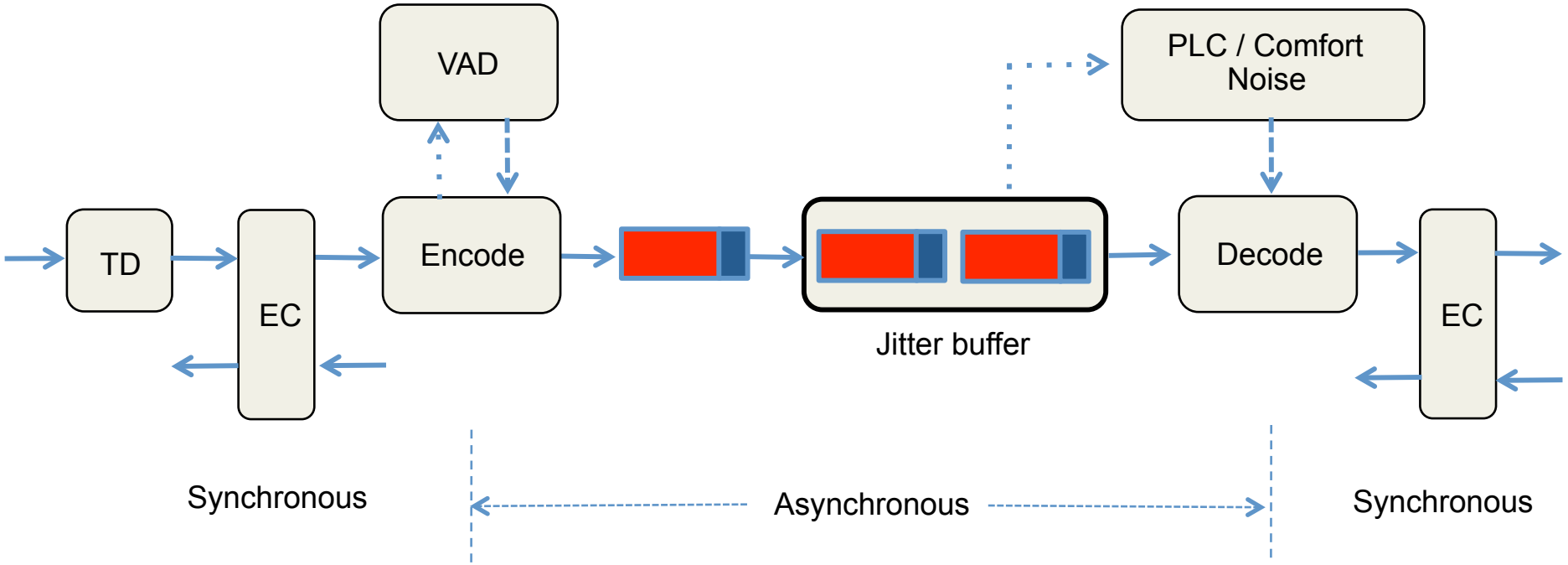
Voice transmission



VoIP: Messaging vs. transmission

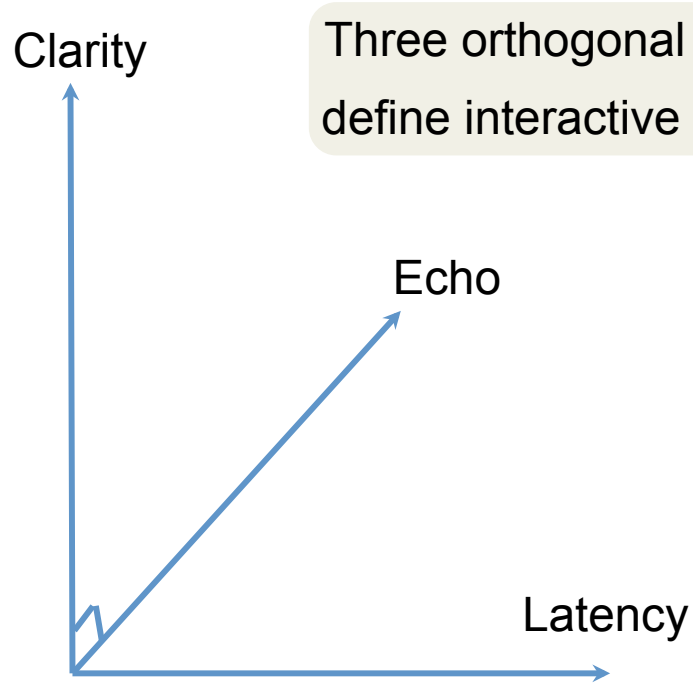


VoIP transmission



Interactive Quality

- Quality
  - Clarity, latency, echo
- Clarity
  - More than intelligibility
  - "ease of use"
  - Factors incl. dist, noise, freq resp, loudness
  - Scale of barely intelligible through 'holographic'



Three orthogonal components define interactive audio quality

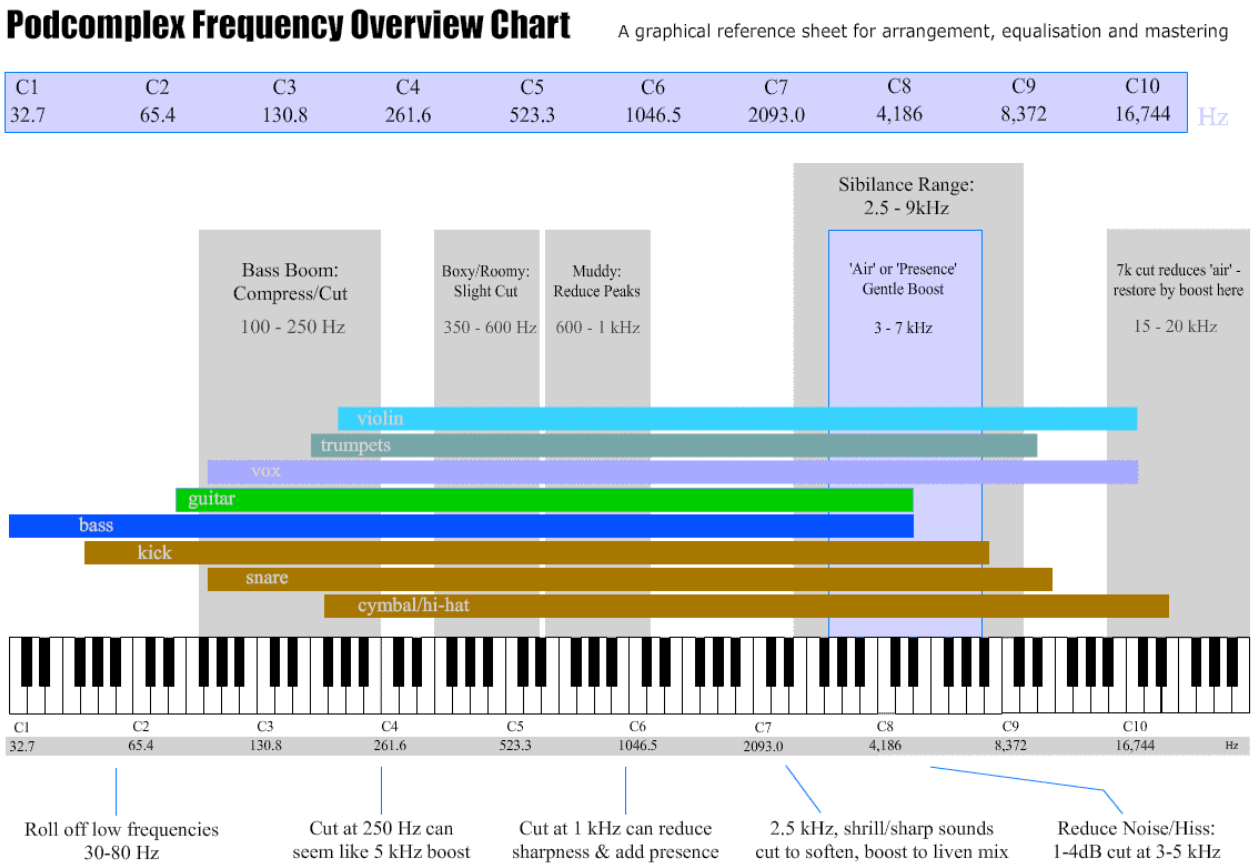
Audio Transmission

Nomenclature	Sampling rate	Usable bandwidth
Narrowband	8 kHz	200 to 3400 Hz
Wideband	16 kHz	50 to 7000 Hz
Super wideband	32 kHz	50 to 14,000 Hz
Fullband	44.1 kHz and up	20 to 20,000 Hz

Useful comparisons: AM radio is limited to 5000 Hz audio
FM radio is limited to 15,000 Hz audio
CD is limited to 20,000 Hz audio

Speed of sound in air: 343 m/s (approx 3 ms/m)

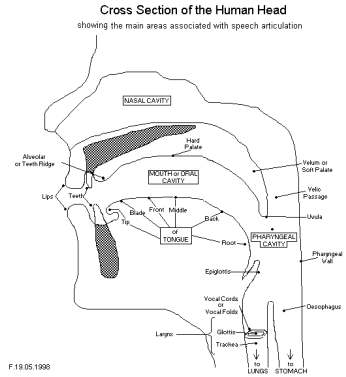
Audio frequencies



http://www.podcomplex.com/images/podcomplex-frequency-overview-chart.gif

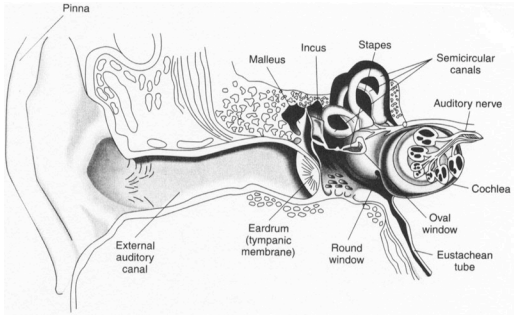
Lossy Compression 101

- Source model based coding
  - Parameterizes source excitation, pitch and formants (a,e,i,o,u)
  - Generally tied to human speech production mechanisms, with limited support for auditory perceptual weighting
  - e.g. G.728, G.729



http://www.sungwh.freeserve.co.uk/sapienti/phon/headxsec.gif

- Perceptual audio coding
  - Uses principals of psychoacoustics and the human auditory system to dynamically assign the most bits to temporal and frequency characteristics most likely to be heard
  - e.g. MP3, AAC
  - Does an MP3 sound ok to a dog?



http://www.skidmore.edu/~hfoley/images/AuditorySystem.jpg

Subjective Testing

- MOS is both a method and metric for subjective quality scoring based on a five point rating system:

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible, but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

- Compressed 4.5 – 5 range makes MOS not suitable for wideband+ quality determination
- MUSHRA (**MUltiple Stimuli with Hidden Reference and Anchor)** with 0-100 scale and more compact statistical requirements better suited

Application Drivers

Application	Channels	Bandwidth	End to end Latency	Allowable complexity	Allowable bit-rate
Speech	**1** - 2	NB - WB	<150 ms	Low	< 64 kbps
Conference	1 - **2**	NB - SWB	Activity driven	Medium	< 128 kbps
Telepresence	2+	SWB - FB	Activity driven	High	< 512 kbps
Gaming	2+	SWB - FB	<150 ms	High	< 320 kbps
Interactive music	2	SWB - FB	< 25 ms	Medium	< 256 kbps

Content: even traditional phone calls handle signal types other than speech (e.g. music-on-hold), as a baseline we must assume non-specific audio content

Other useful features: packet loss concealment, quality and bandwidth layering, joint multi-channel encoding

Narrowband matrix (8 kHz fs)

Codec	Bit rate (kbps)	Look ahead (ms)	Frame size (ms)	PSQM (zero impair)	DTX	PLC
G.711	64	0	Arbitr.	4.45	Appendix II	Appendix I
G.723.1	5.3, 6.3	7.5	30	3.6, 3.9 (MOS)	Yes	Yes
G.728	16	0	0.562	3.6 (MOS)		
G.729AB	8	5	10	4.04	Yes	Yes
AMR	4.75 – 12.2	5	20	4.14	Yes	Yes
GSM-EFR	12.2	0	20 or 30			Yes
iLBC	13.33, 15.2	0	20 or 30	4.14 (15.2)		Yes

Sources: http://en.wikipedia.org/wiki/Comparison_of_audio_formats,
Cable Labs PKT-SP-CODEC-MEDIA-I08-100120

Wideband +

Codec	Sample rate (kHz)	Bit rate (kbps)	Algorithm latency (ms)	Comp Cmplx	# Chan	PLC
G.711.1	8, 16	64, 80 (8 kHz) 80, 96 (16 kHz)	11.875		1	
G.718	8, 16 (extens.)	8 - 32	42.875 – 43.875 (20 ms frames)		1	Yes
G.719	48	32 - 64	40 (20 ms frames)	18 FP-MIPS	1, MC (MP4)	
G.722	16	64	4	10 MIPS		No
G.722.1(C)	16, 32 (c)	24, 32, 48 (32)	40 (20 ms frames)	10 WMOPS		Yes
G.722.2 (AMR-WB)	16	6.6 – 23.85	25	38 WMOPS	1, MC (MP4)	Yes
G.729.1	8, 16	8 - 32	48.9375			Yes
Siren	16 - 48	16 (m) – 128 (s)	40 (20 ms frames)		1 or 2	
Speex	8 - 32	2 - 44	30 NB, 34 WB		1, 2 opt.	Yes
AAC-ELD	? - 48?	24 - 64	15 (64) – 32 (24)		1+	Yes

Summary

- Goal 1: set codec application space -> define parameters of interest
- Goal 2: survey current codecs and works-in-progress
- Goal 3: define benchmark tools and performance goals
- Goal 4: qualify codecs, make choice(s)