Just The Essentials of Voice Encoding over IP

Technology White Paper

Serge-Paul Carrasco

Fundamentals of Digital Signal Processing

Basic Signal Processing Theory

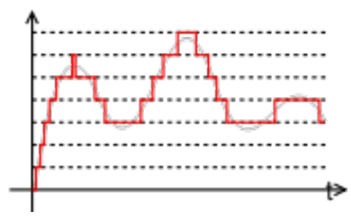
- Digitized media: analog-to-digital conversion
- Synthesized media: digitally generated signal

Sampling

- Samples: measurements of the analog signal (which is a continuous-time signal).
- Sampling: providing a discrete-time representation of the analog signal at uniform intervals to express the measurements (e.g. the samples).
- Sampling frequency:
  - Given by the Nyquist theorem: the minimum sampling rate for a signal with a frequency in [o,w] range is 2w.

Quantization

- Process of approximating a continuous range of values by a relatively small set of integers.



Quantized Signal

- Quantization introduces an error that cannot be recovered.

Analog to Digital Conversion (A/D)

- Provide sampling and quantization for a specific number of bits.
- Process:
  - Define the range of the analog signal:
    - Example: S(i) = [-1, 1]
  - Define the number of bits of the digital signal:
    - Example: 3 bits, possible quantization vectors are: Q(j)={000,001,010,011,100,101,110,111}
  - Map Q(j) to S(i) scale:
    - Each vector is mapped to an analog value
  - For any given S(i), calculate its distance to all Q(j). The shortest distance will give the right Q(j).

Copyright © 2003-2010 by Serge-Paul Carrasco. All rights reserved.

- Pulse Code Modulation (PCM):
    - A/D conversion for phone call sampled at 8,000 per seconds over 8 bits producing a 64 kb/s signal (DS0)
- Variations to PCM:
    - Differential (or Delta) PCM (DPCM):
        - Encodes the PCM values as differences between the current and the previous. Reduce the number of bits per sample
    - Adaptive PCM:
        - Varies the size of the quantization (because the voice signal changes). Further reduce bandwidth need for as given signal to noise ratio (S/N)

Linear Prediction Coding (LPC)

- Mathematical combination of previous samples to construct a predicted value that attempts to approach the next input sample.

Companding

- Compressing and expanding (with a logarithmic scale) the voice signal to optimize it (small errors for a small amplitude and large errors for a large amplitude):
    - A-Law is the algorithm for Europe for DS0/PCM
    - Mu-Law is the algorithm for NA and Japan for DS0/PCM

Type	**Signal (hz)**	**Signal Samples (kHz)**	**Quantization (bits)**	**Transmission (kbit/s)**	**Applications**
Telephone speech	300-3,400	8	12 or 13	96 or 104	PSTN, Digital Cellular
Wideband speech & audio	50-7,000	16	14 or 15	224 or 240	Video and audio conferencing, FM radio
High quality Speech & audio	30-15,000 20-20,000 10-22,000	32 44.1 48	16 16 Up to 24	512 706 1,152	TV audio Audio CD player Professional audio

Audio Signals Parameters

Copyright © 2003-2010 by Serge-Paul Carrasco. All rights reserved.

VoIP Quality

Network Parameters Affecting VoIP Quality

- Jitter:
    - Packets within one talkspurt (time period of speech sounds unbroken by silent moments) arrive at their destinations with varying delays
    - Compensation methods:
        - Jitter-buffer talkspurt by talkspurt
- Delay:
    - Caused by network transmission, jitter buffer and coding/decoding processes
- Packet loss:
    - Wireline networks: excessive network congestion
    - Wireless networks: bit errors introduced on radio link
    - Compensation:
        - Packet loss concealment (PLC):
            - Lost packets are compensated at the receiver side based on coding information from previously received packets
            - Forward error correction (FEC):
                - Redundant transmission of the coded speech data
            - Low bit rate redundancy (LBR):
                - Redundant version of the coded speech data at lower bit rate
        - Both FEC and LBR increase the network delay
    - **Packet loss is the most important quality factor for VoIP**
- Packet size:
    - Limit the number of packet sent to limit the delay incur by the processing of the packet header
    - But larger packets lead to an increase in the transmission delay
- Bit errors:
    - Due to inference/errors with signal transmission
- Noise:
    - Line noise
    - Signal-correlated noise
    - Background noise
- Linear distortion

User Interfaces Parameters Affecting VoIP Quality

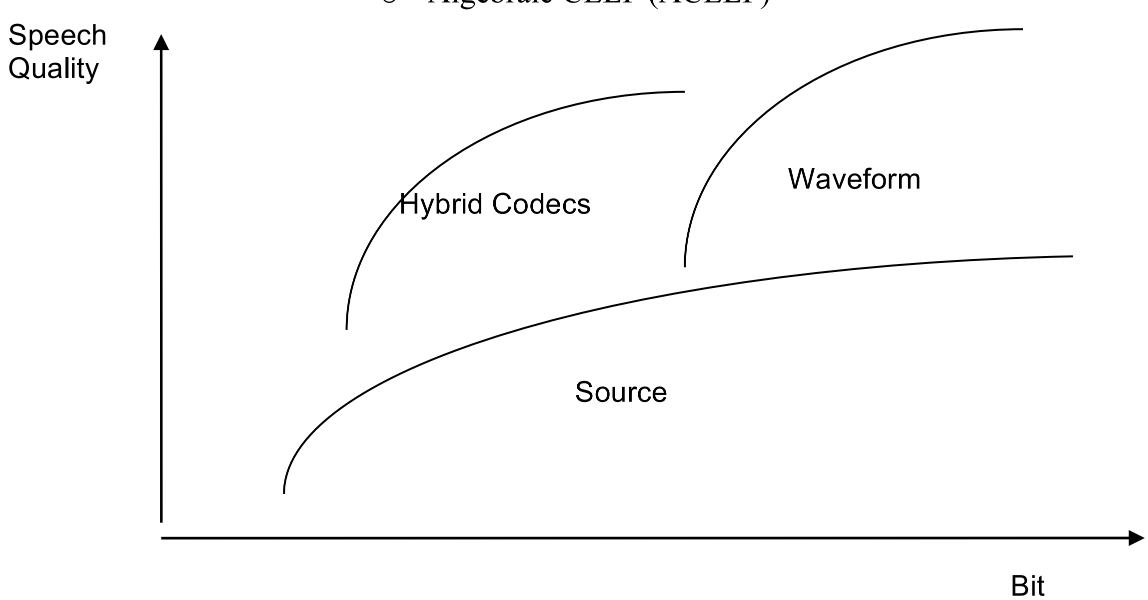
- Talker echo
- Listener echo
    - Compensation:
        - Echo cancellers
- Loudness of the speech
- Background noise

Copyright © 2003-2010 by Serge-Paul Carrasco. All rights reserved.

Speech Codecs

Speech Codecs Types

- Waveform codecs or temporal speech codecs:
  - Based on quantization of the actual speech waveform (e.g. signal)
  - High speech quality for high-bit rates:
    - Efficient from 20-40 kb/s
    - Degrades around 16 kb/s
- Source codecs or vocoders or parametric codecs:
  - Based on a mathematical speech production model
  - Low speech quality (sounds seem synthetic) codecs
  - Not used in public networks
- Hybrid codecs:
  - Combination of waveform and source codecs:
    - Waveform matching with knowledge of speech production
  - High speech quality for low bit rates
  - Types:
    - Analysis by synthesis (ABS):
      - Based on linear prediction coding
    - Variations of ABS codecs:
      - Regular-Pulse Excited (RPE)
      - Codebook-excited linear predictive (CLEP)
        - Algebraic CELP (ACELP)



Codecs Types

Codec Standards

Copyright © 2003-2010 by Serge-Paul Carrasco. All rights reserved.

- Narrowband codecs (300-3,400 hz):
  - PSTN:
    - G.711 (Waveform/PCM) (Reference codec for NB):
      - G.711 encodes a 12-bits linearly quantized signal into 8 bits
    - G.721 (Waveform/ADPCM)
    - G.726 (Waveform/ADPCM)
  - Mobile/GSM:
    - Full rate (FR)
    - Enhanced FR (EFR)
  - Packet networks:
    - G.711
    - G.723
    - G.723.1
    - G.729
  - Specifically for VoIP:
    - iLBC (GIPS open source codec)
    - Adaptive Multirate Codec- Narrowband (AMC-NB)
- Wideband codecs (cover 50-7,000 hz):
  - PSTN:
    - G.722 (Reference codec for WB)
    - G.722.1
    - WB extensions to G.729
  - Mobile/GSM:
    - WB-AMR (or G.722.2)

	G.711	G.726	G.729	G.723.1	G.722	G.722.2
Date of approbation	1972	1990	1995 & 1999	1995	1999	2002
Bit rate (kb/s)	64	16/24/32/40	8/6.4/11.8	6.3/5.3	32/24	Multiples
Speech Quality (MOS)	4.2	4.0	4.0	4.0		
Coder Type	Waveform PCM	Waveform ADPCM	Hybrid ABS LD-CELP	Hybrid ABS ACELP	Hybrid LCTC	Hybrid ACELP

Major ITU-T Codecs

Quality of Speech Codecs

Copyright © 2003-2010 by Serge-Paul Carrasco. All rights reserved.

- Quality of the speech reproduced
- Delay introduced by the coder algorithm
- The complexity of the coder that will result in additional processing delay
- The behavior of the coder for music, modem and DTMF
- Tandeming properties:
    - Number of times voice can be encoded and decoded

- Mean opinion score (MOS):
    - Mark from 0 to 5 given to an audio sample by a group of listeners

Telephony and Multimedia Networks

Telephony Networks

- PSTN:
    - SS7/ISDN: Signaling for PSTN
- Packet media transport networks:
    - H.323: Initial signaling for VoIP from ITU
    - MGCP/H.248: Softswitch architecture from Bellcore/MEGACO from IETF
    - SIP: IETF signaling for VoIP
    - IP Multimedia Subsystem (IMS): Integrate IP multimedia to wireless networks
- VoIP speech payload:
    - Transported over RTP/UDP/IP

Copyright © 2003-2010 by Serge-Paul Carrasco. All rights reserved.