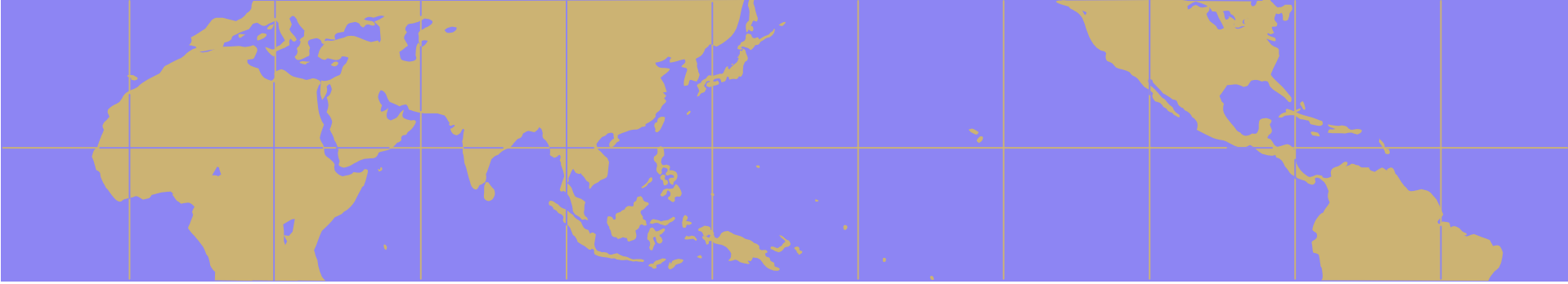
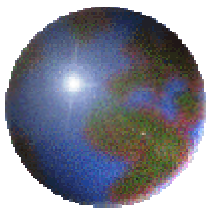


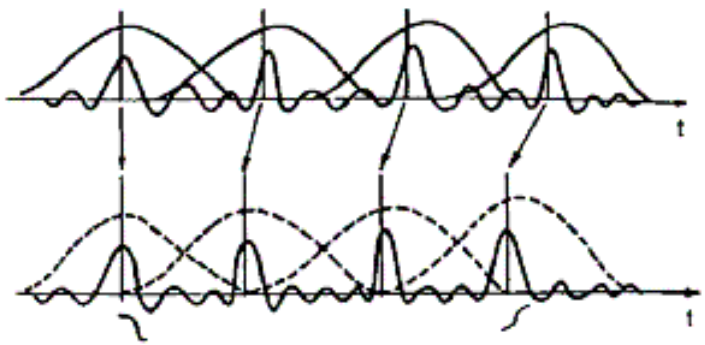


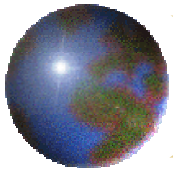
Speech Coding Standards

Antti Kiviluoto

Antti.Kiviluoto@hut.fi

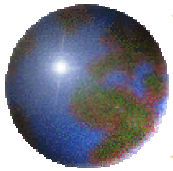
Speech coding standards –methods that enable voice communication





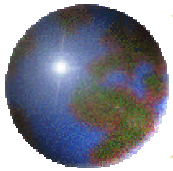
Presentation

- Intro
- Speech Analysis - Synthesis & Linear Prediction
- Speech coding standards
- Subband & Transform Coders
- Applications for synthetic speech
- Summary



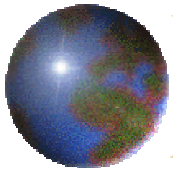
Intro

- Interest towards speech coding & standardization:
  - World wide growth in communication networks
  - Emergence of new multimedia applications
  - Advances in Very Large-Scale Integration (VLSI) devices
- Standardization
  - International Telecommunications Union (ITU)
  - European Telecom. Standards Institute (ETSI)
  - International Standards Organization (ISO)
  - Telecommunication Industry Association (TIA), NA
  - R&D Center for Radio systems (RCR), Japan



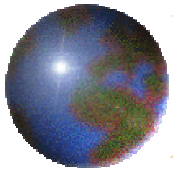
Intro

- Signal models are based on Linear Predictive Coding (LPC)
  - Open-loop linear prediction (LP)
  - Closed-loop/Analysis-by-synthesis LP
  - Subband & transform coding algorithms
- Quality standards for SCA
  - Mean Opinion Score (MOS)
  - Diagnostic Rhyme Test (DRT)
  - Diagnostic Acceptability Measure (DAM)



Speech Analysis - Synthesis & Linear Prediction

- Speech is produced by the interaction of the vocal tract with the vocal chords in the glottis
- Engineering model:
  - Vocal tract - time-varying digital filter
  - Filter parameters estimated by using LP algorithms

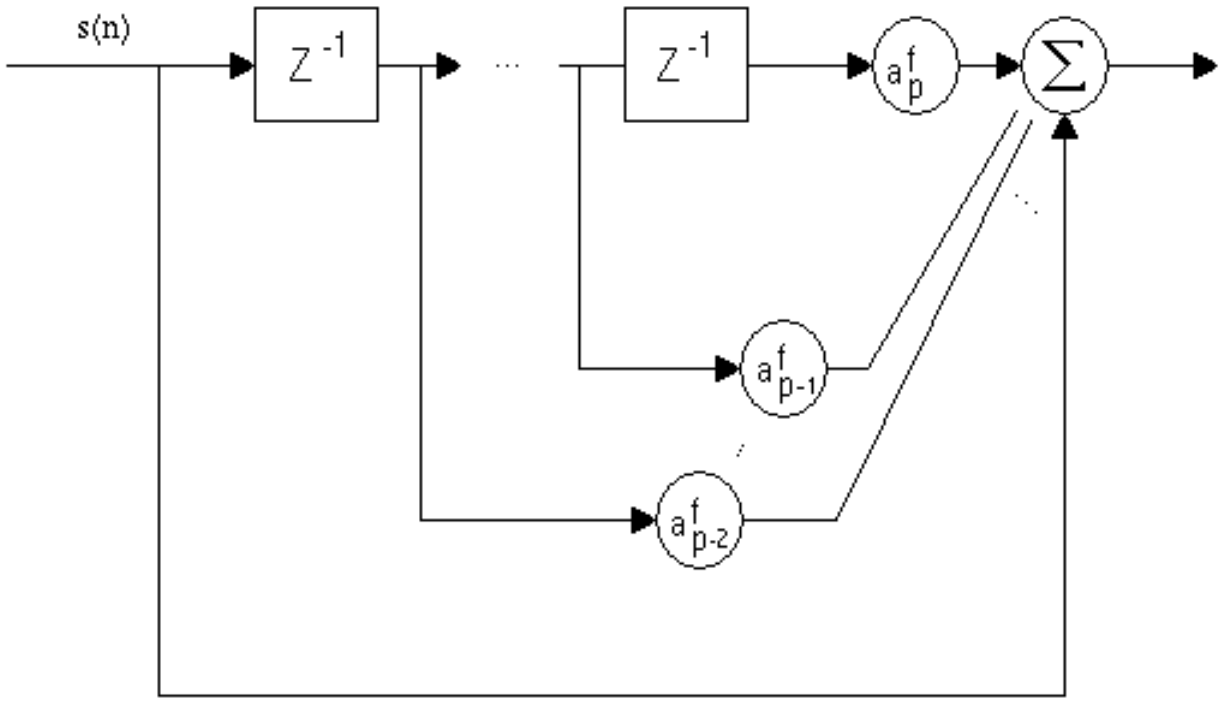


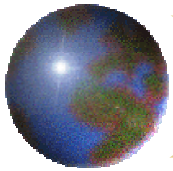
Speech Analysis - Synthesis & Linear Prediction

- Linear Prediction
  - Introduced in late 1960's
  - Widely used in speech signal processing
  - Number of available techniques since introduction
  - Basic principle of prediction still the same
  - Most recent sample predicted by a linear combination of past samples
  - LP analysis done by Finite-length Impulse Response (FIR) digital filter

Speech Analysis - Synthesis & Linear Prediction

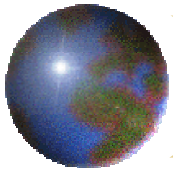
- Linear prediction analysis





Speech Analysis - Synthesis & Linear Prediction

- Open-loop linear prediction (waveform)
  - Low complexity codecs
  - Good quality with rates above 16 Kbit/s
  - Pulse Code Modulation (PCM)
    - simplest form of waveform codecs
    - Non-linear quantization
    - quantizing the speech signal directly
  - Adaptive Differential PCM (ADPCM)
    - quantize the difference between the speech signal and a prediction

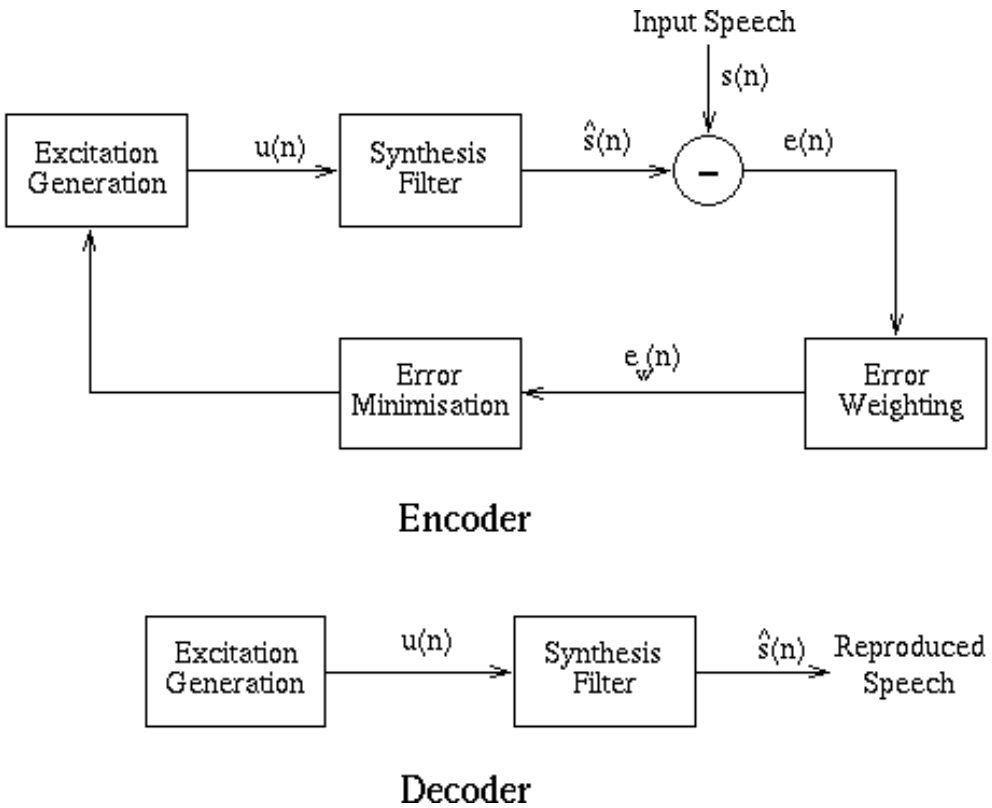


Speech Analysis - Synthesis & Linear Prediction

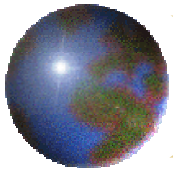
- Closed-loop/Analysis-by-synthesis LP
  - introduced in 1982
  - Intelligible speech at 2.4 kbits/s and below
  - Frame size usually 20 ms
  - Cannot provide “natural” sounding speech at any bit rate
  - Multi-Pulse Excitation (MPE) codec
  - Regular-Pulse Excitation (RPE) codec
  - Code-Excitation Linear Predictive (CELP) codec
  - Vector Sum Excited Linear Prediction (VSELP)

Speech Analysis - Synthesis & Linear Prediction

- AbS Codec Structure

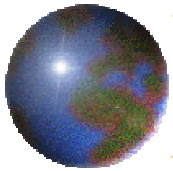


- Splitting the input speech to be coded into frames
- Parameters determined for Synthesis filter
- Excitation determined to the filter
- Encoder transmits information representing the filter parameters and the excitation to the decoder
- at the decoder the given excitation is passed through the synthesis filter to give the reconstructed speech



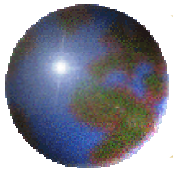
Speech Analysis - Synthesis & Linear Prediction

- Short-Term Prediction (STP)
  - Captures the formant structure of short-term speech spectrum
- Long-Term Prediction (LTP)
  - Captures the long-term correlation in the speech signal
  - Representing the periodicity in speech



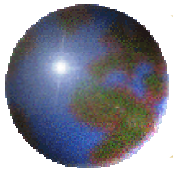
Speech Analysis - Synthesis & Linear Prediction

- Multi-Pulse Excitation (MPE) codec
  - Skyphone standard of British Telecom International, used in passenger communications in aircraft
  - Excitation is a sequence of multiple nonuniformly spaced pulses
  - Uses both short- and long-term prediction
- Regular-Pulse Excitation (RPE) codec
  - Used in GSM standard
  - Excitation is a sequence of multiple uniformly spaced pulses
  - Uses only long-term prediction



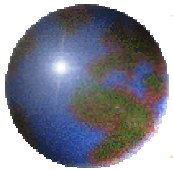
Speech Analysis - Synthesis & Linear Prediction

- Code- Excitation Linear Predictive (CELP) codec
  - Codebook contains excitation vector
  - Uses long- and short-term prediction



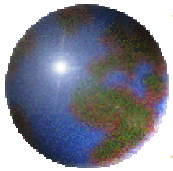
Speech Analysis - Synthesis & Linear Prediction

- Quality standards for SCA
  - Evaluation of speech quality and intelligibility
- Mean Opinion Score (MOS)
  - Quite common in standardization tests
  - Five-level quality scale, 5 (excellent) – 1 (bad)
    - 4.0-4.5: network or toll quality
    - 3.5-4.0: communications quality (cellular grade)
    - 2.5-3.5: synthetic quality



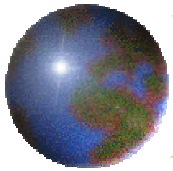
Speech coding standards

- ITU G.726 and G.727 ADPCM Coders
  - Quantize the difference between the speech (current) signal and a predicted speech samples
  - Prediction parameters are obtained by backward estimation, from already quantized data
  - Algorithm operates at 16, 24, 32, 40 and 64 Kbit/s
  - G.726: individually optimized quantizers
  - G.727: embedded quantizers - developed for packet network applications (switch rates)
  - MOS for 32 Kbit/s: 4.1
  - Complexity: 2 MIPS



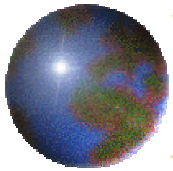
Speech coding standards

- Inmarsat-B Adaptive Predictive Vocoder
  - The International Mobile Satellite B standard
  - Uses ADPCM coder with LTP and 6th order STP
  - Algorithm operates at 9.6 and 12.8 Kbit/s
  - MOS for 9.6 (12.8) Kbit/s: 3.1 (3.4)
  - Complexity: 10 MIPS



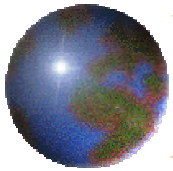
Speech coding standards

- LPC-10(e) Federal Standard 1015
  - Introduced by Department of Defense in 1976
  - Secure communications
  - Uses 10^{th} order predictor
  - Algorithm operates at 2.4 Kbit/s
  - MOS for 2.4 Kbit/s: 2.3
  - Complexity: 7 MIPS



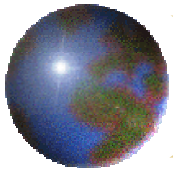
Speech coding standards

- Mixed Excitation Linear Prediction (MELP) codec
  - Introduced by U.S. government in 1996
  - Development motivated by voicing errors in LPC-10
  - Algorithmic delay estimated 122.5 ms
  - Algorithm operates at 2.4 Kbit/s
  - MOS for 2.4 Kbit/s: 3.2
  - Complexity: 40 MIPS



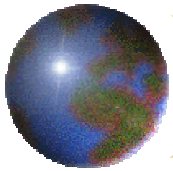
Speech coding standards

- ETSI full-Rate GSM 6.10
  - Pan-European digital cellular standard
  - Uses RPE coder with LTP
  - Algorithm operates at 13.0 Kbit/s
  - MOS for 13.0 Kbit/s: 3.5-3.9
  - Complexity: 5-6 MIPS



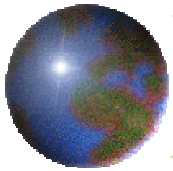
Speech coding standards

- FS-1016
  - Used by Department of Defense (DoD) for 3rd-generation secure telephone unit (STU-III)
  - Developed jointly by DoD and AT&T Bell Laboratories
  - Uses CELP algorithm
  - excitation is formed by combining vectors from LTP and stochastic codebook
  - Algorithm operates at 4.8 Kbit/s
  - MOS for 4.8 Kbit/s: 3.2
  - Complexity: 16 MIPS



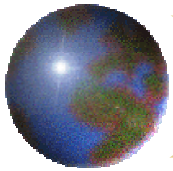
Speech coding standards

- Vector Sum Excited LP (VSELP) Standards
  - Algorithm embedded to 3 digital cellular standards
    - TIA IS-54 (Japan & North America)
    - ETSI 6.20 GSM half-rate (Europe)
  - Highly structured codebooks
    - Reduceses computational complexity
    - Increases robustness to channel errors
  - Algorithm operates at 2.4 Kbit/s
  - MOS for 6.3 (8.0) Kbit/s VSELP: 3.4 (3.5)
  - Complexity: 14 MIPS



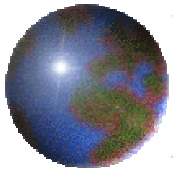
Speech coding standards

- ITU G.728 Low-Delay CELP
  - Uses low-delay CELP coder
    - Short frames
    - Short exication vectors
    - No LTP and the order of STP is increased to 50
  - Speech quality equivalent or better than G.726
  - Algorithm operates at 16.0 Kbit/s
  - MOS for 16.0 Kbit/s: 3.4
  - Complexity: 30 MIPS



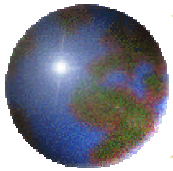
Speech coding standards

- IS-96
  - Used in Code Division Multiple Access (CDMA), standard for cellular communications
  - Uses CELP algorithm with STP
  - Bit rate is variable, determined by speech activity
  - Lower rates achieved by allocating fewer bits to LP parameters
  - Algorithm operates at 1.2, 2.4, 4.8 and 9.6 Kbit/s
  - MOS for 9.6 Kbit/s: 3.3
  - Complexity: 15 MIPS



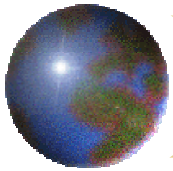
Speech coding standards

- ITU G.729 and G.729A CS-ACELP
  - Designed for wireless and multimedia network applications
  - Uses Conjugate Structure Algebraic CELP
    - Low-delay algorithm (15 ms)
    - Frame size 10 ms
    - Uses 2 codebooks
  - G.729 Annex has lower complexity and quality
  - Algorithms are interoperable
  - Algorithm operates at 8.0 Kbit/s
  - MOS for 8.0 Kbit/s: G.729=4, G.729A=3.8
  - Complexity: G.729=20 MIPS, G.729A=11 MIPS



Speech coding standards

- ITU G.723.1 MP-MLQ/ACELP
  - Speech coder for audio and videoconferencing over public telephone networks
  - Part of ITU H.323 and H.324 standards
  - Also recommended to be default audio codec for voice of the network (VOIP)
  - Standard is dual-rate with 2 excitation schemes
    - Multipulse maximum likelihood quantization for 6.3 Kbit/s
    - ACELP for 5.3 Kbit/s
  - MOS for 5.3 (6.3) Kbit/s: 3.7 (4.0)
  - Complexity: 16 MIPS

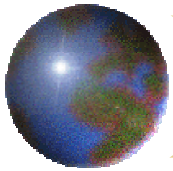


Subband & Transform Coders

- Subband Coder
  - The signal is divided into frequency subbands
  - Coder exploits the statistics of the signal and encodes each band using a different number of bits
  - E.G lower-frequency bands are allotted with more bits
    - Preserve critical pitch and formant information
  - ITU G.722, standard for ISDN teleconferencing
    - Low frequency subband quantized with 48 Kbit/s
    - High-frequency with 16 Kbit/s
  - MOS for 64 Kbit/s: 4.1
  - Complexity: 5 MIPS

Subband & Transform Coders

- Sinusoidal Transform Coding
  - Speech is represented by a linear combination of sinusoids
  - Opportunity to reduce the bit rate
    - Speech is typically periodic
    - Lots of pauses
  - Considered for wideband and high-fidelity app.
  - MOS for 2.4 (4.8) Kbit/s: 2.9 (3.5)
  - Complexity: 13 MIPS

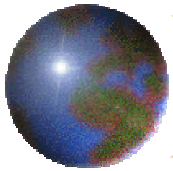


Applications for synthetic speech

- Telecommunications and Multimedia
- For the Blind
- For the Deafened and Vocally Handicapped
- Educational
- Samples...

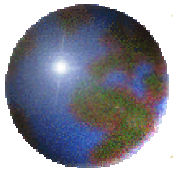
Summary

Algorithm	Bit Rate (Kbit/s)	MOS	Complexity (MIPS)	Frame Size (ms)
PCM G.711	64	4.3	0.01	0
ADPCM G.726	32	4.1	2	0.125
SBC G.722	48/56/64	4.1	5	0.125
LD-CELP G.728	16	4.0	30	0.625
CS-ACELP(-A) G.729	8	4.0 (3.8)	20 (11)	10
MPC-MLQ G.723.1	6.3/5.3	4.0/3.7	11	10
GSM HR VSELP	6.3	3.4	14	20
IS-54 VSELP	8	3.5	14	20
IS-96 QCELP	1.2/2.4/4.8/9.6	3.3	15	20
Inmarsat-B APC	9.6/12.8	3.1/3.4	10	20
MELP	2.4	3.2	40	22.5
FS 1016 – CELP	4.8	3.2	16	30



Summary

- Linear prediction widely used in speech coding
- Speech coding research has come a long way in the last 10 years
- Need for speech codecs from cellular phones to multimedia systems



The End

- Time for questions