

Speech and Audio Coding Based on Temporal Contexts in Sub-bands

Workshop on Temporal Dynamics in Speech and Hearing
Antwerp, Belgium, August 26, 2007

Harinath Garudadri hgarudadri@qualcomm.com

Petr Motlicek petr.motlicek@idiap.ch

Sriram Ganapathy sriram.ganapathy@idiap.ch

Hynek Hermansky hynek.hermansky@idiap.ch



Problem

- Vcoders used in voice services do not handle mixed content well
- Audio codecs used in multimedia do not compress speech well

Notable Initiatives

- AMR-WB+, AAC-ELD, MPEG “Speech + Audio Explorations”, G.729.1, G.VBR, ...

Applications: Data rich, 3G multimedia services

- Broadcast, download, store and push, store and pull,
- Latency requirements are not stringent

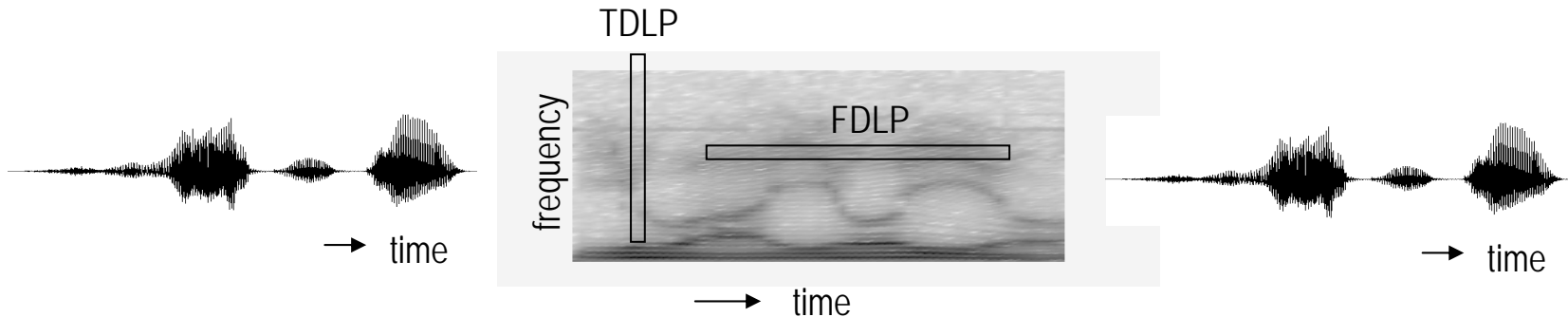


Question: What can we do if we had “all the time” in the world?

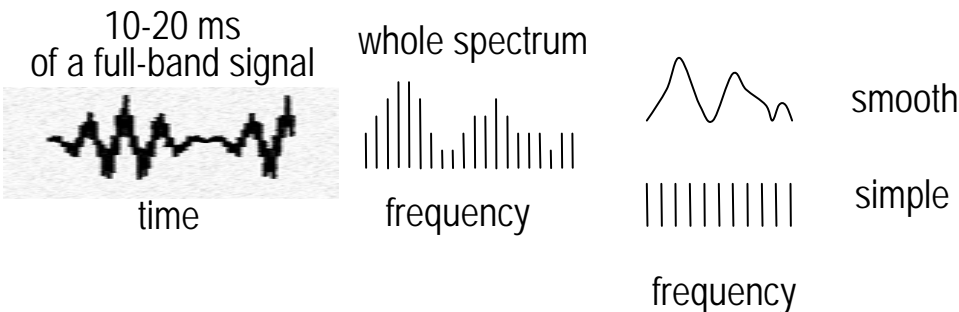


Modeling Spectral Dynamics in Sub-bands

signal □ → analyze → select and quantize → transmit → reconstruct → signal □



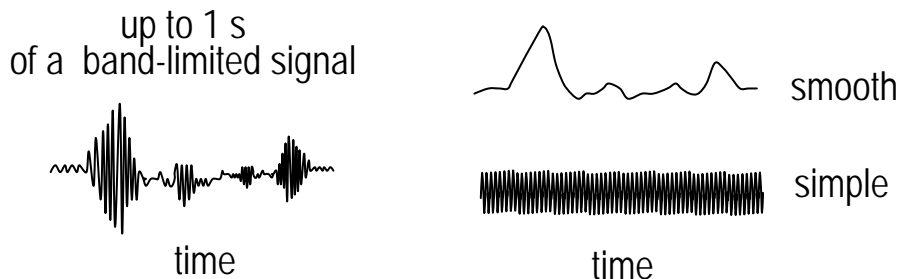
Time-domain (conventional) linear prediction



Attempt to model of speech production
(source-filter linear model of speech production)

10-20 ms algorithmic delay
Efficient for coding speech
Complete signal as a concatenation of short segments
Dropouts cause loss of the segment of the signal

Frequency-domain linear prediction



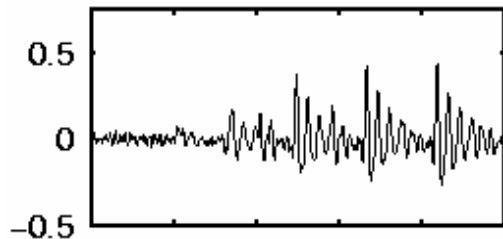
Attempt to model auditory perception
(frequency selectivity of hearing)

1000 ms algorithmic delay
Suitable also for non-speech signals
Complete signal as a sum of sub-band components
Dropouts cause loss of a parts of the spectrum

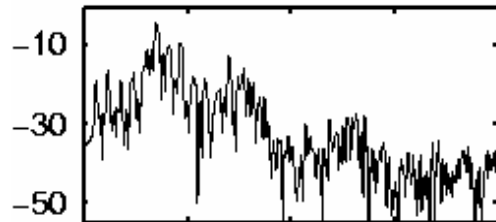
All Pole Model of Spectral Tracks

conventional LP

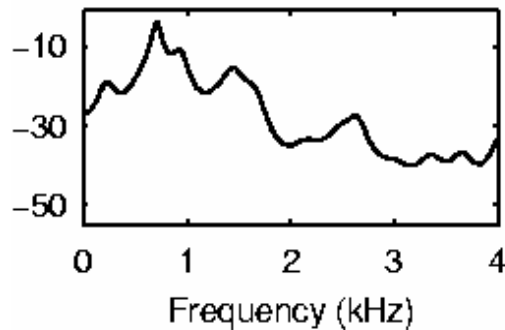
the signal



signal power spectrum

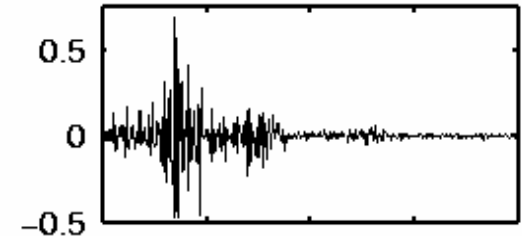


all-pole model of the power spectrum

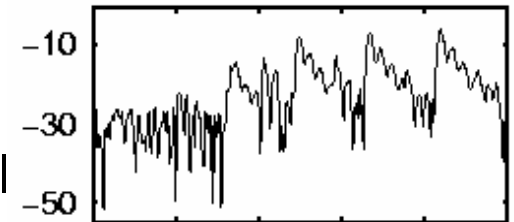


spectral domain LP

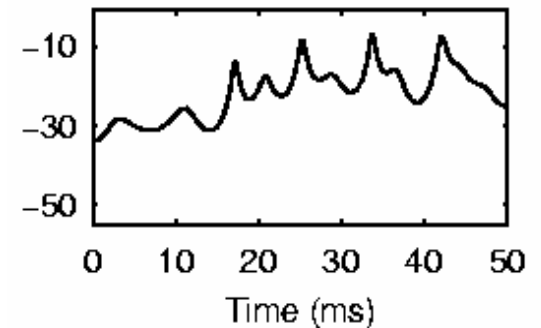
DCT of the signal



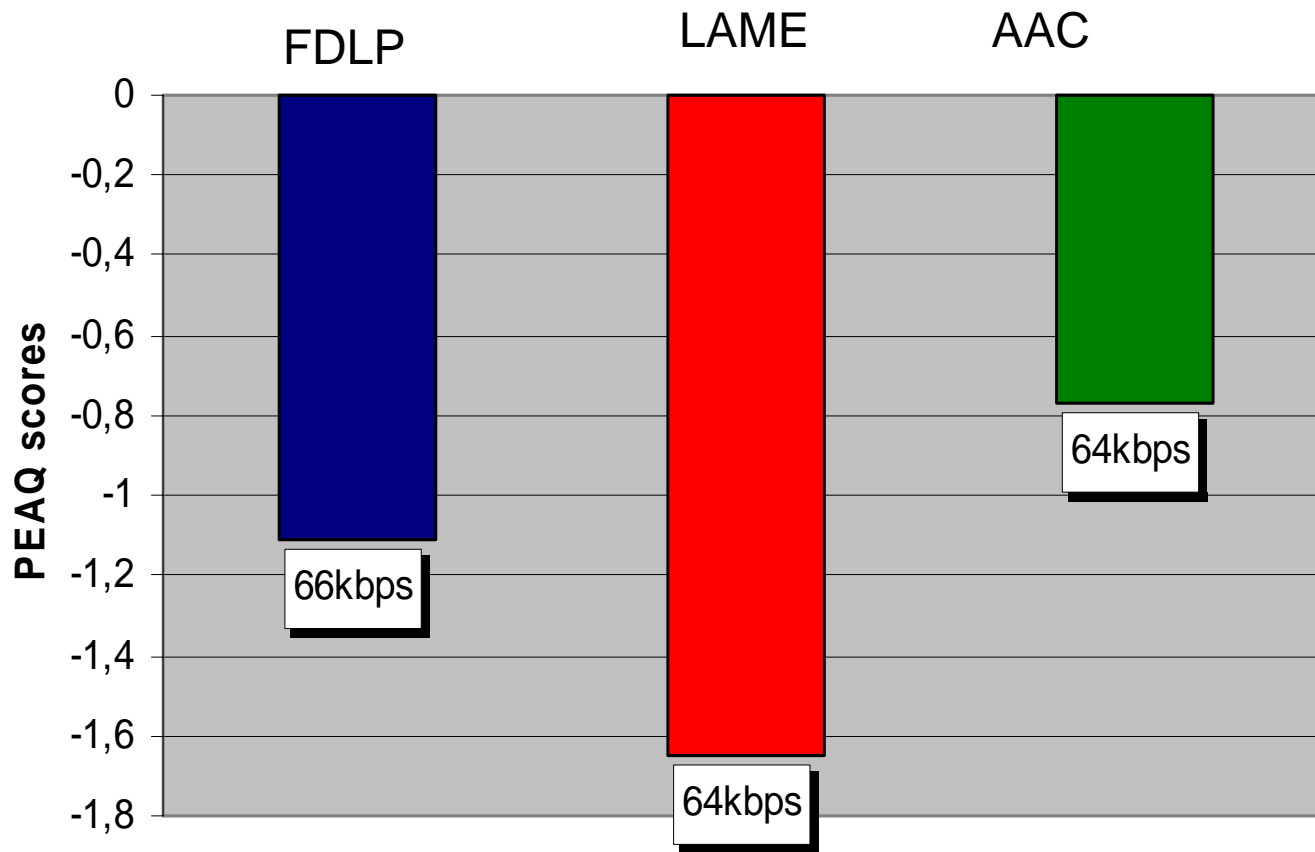
Hilbert envelope of the signal



all-pole model of the Hilbert envelope

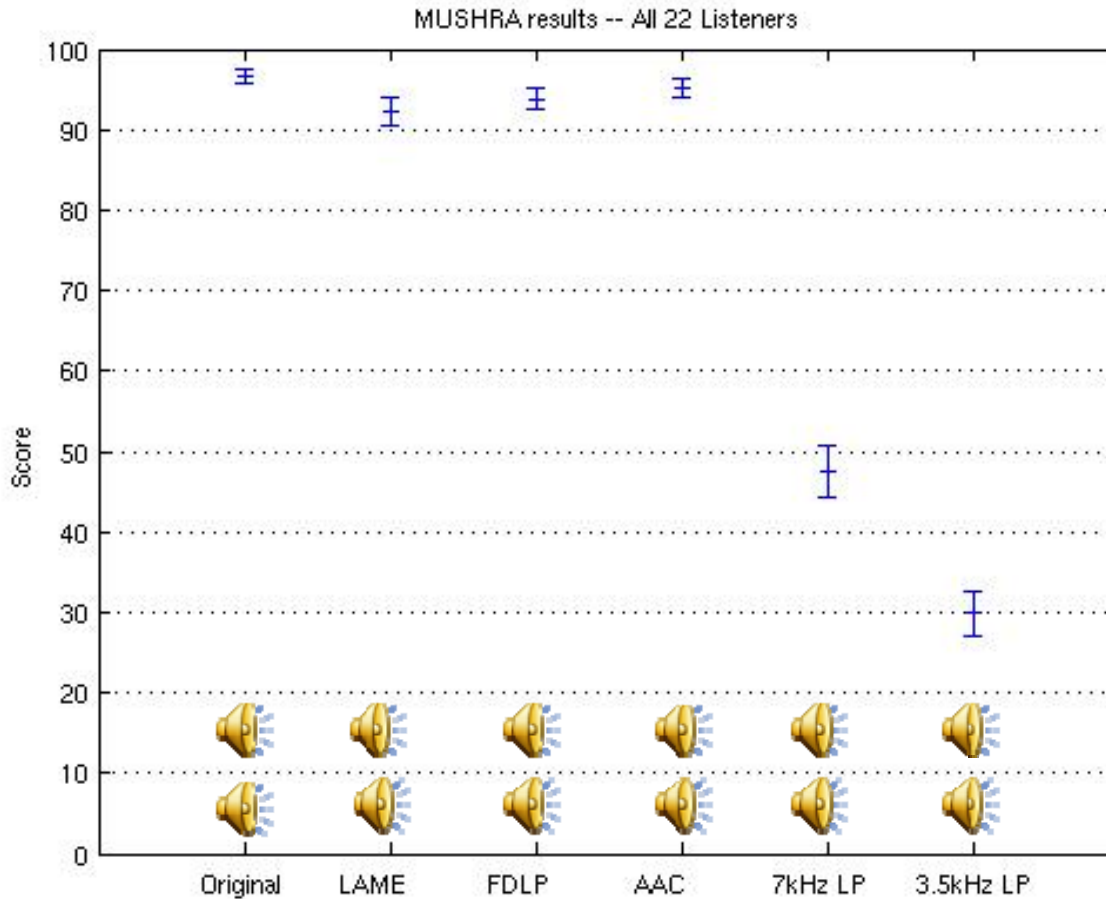


Quality Evaluation



- MPEG Database for “Exploration of Speech and Audio Coding”
- Objective Evaluation: ITU-R BS.1387 PEAQ (27 files)
- LAME = mp3; AAC = aacPlus v1

Quality Evaluation



- Subjective Evaluation: ITU-R BS-1534 MUSHRA (22 listeners)
- Content = 8 files (Music 4 files; Speech + Music 2 files; Speech 2 files)
- LAME = mp3; AAC = aacPlus v1

Latency is not critical in many 3G multimedia services

Frequency Domain Linear Prediction

- Modeling temporal dynamics in sub-bands
- Unified approach for speech, audio and mixed content
- Resilient to dropouts (not discussed in this presentation)

Experiments

- MPEG Database for “Exploration of Speech and Audio Coding”
- Subjective Evaluation: ITU-R BS-1534 MUSHRA (22 listeners)
- Objective Evaluation: ITU-R BS.1387 PEAQ (27 files)

Results

- Audio quality (around 64 kbps) is similar to that of state of the art MPEG codecs