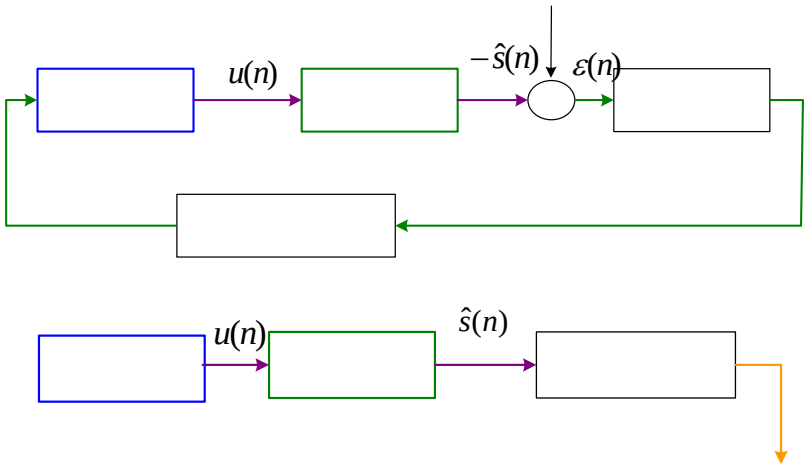
9. CLOSED-LOOP SPEECH CODING TECHNIQUES

Closed-Loop (Analysis-by-Synthesis) LPC Speech Coding

- Analysis-by-Synthesis (AbS) or closed-loop LPC speech coding introduced in 1982 by Atal and Schroeder as a 2400 bits/second standard codec with intelligible quality
- Frame size fixed at usually 20 ms.
- Post-synthesis enhancements
- Transformed prediction coefficients
- Used in next generation federal and commercial standards and basis for the current cellular phone speech codecs including:
  1. Multi-pulse Excitation (MPE)
  2. Regular-Pulse Excitation (RPE) of GMS codecs
  3. Code-Excited Linear Prediction (CELP) codec family (Ex: QCELP)
  4. Vector Sum Excited Linear Prediction (VSELP) codec family: (Ex: TDMA systems using Motorola codecs)
  5. Mixture Excitation Linear Prediction (MELP) codec family.

Analysis-by-Synthesis (AbS) LPC Structure



Short-Term Prediction and Long-Term Prediction AbS Codecs

- STP measures the short-term correlation for capturing vocal tract of speech spectrum.
- LTP captures the long-term correlation (pitch-to-pitch) in the speech spectrum.

Illustration of short-term and long-term correction of speech segments:

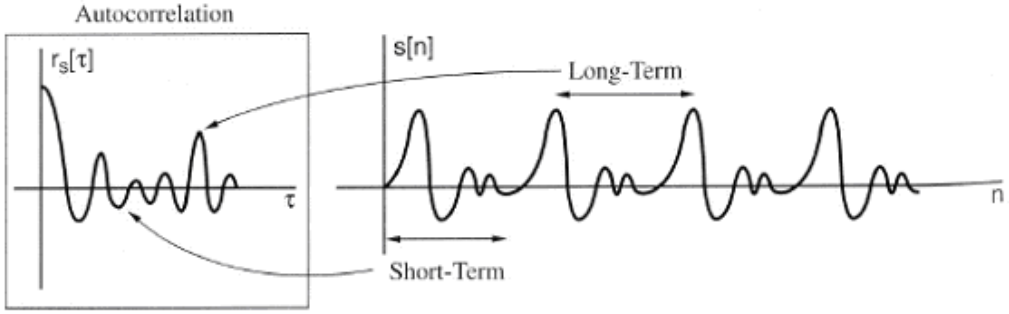
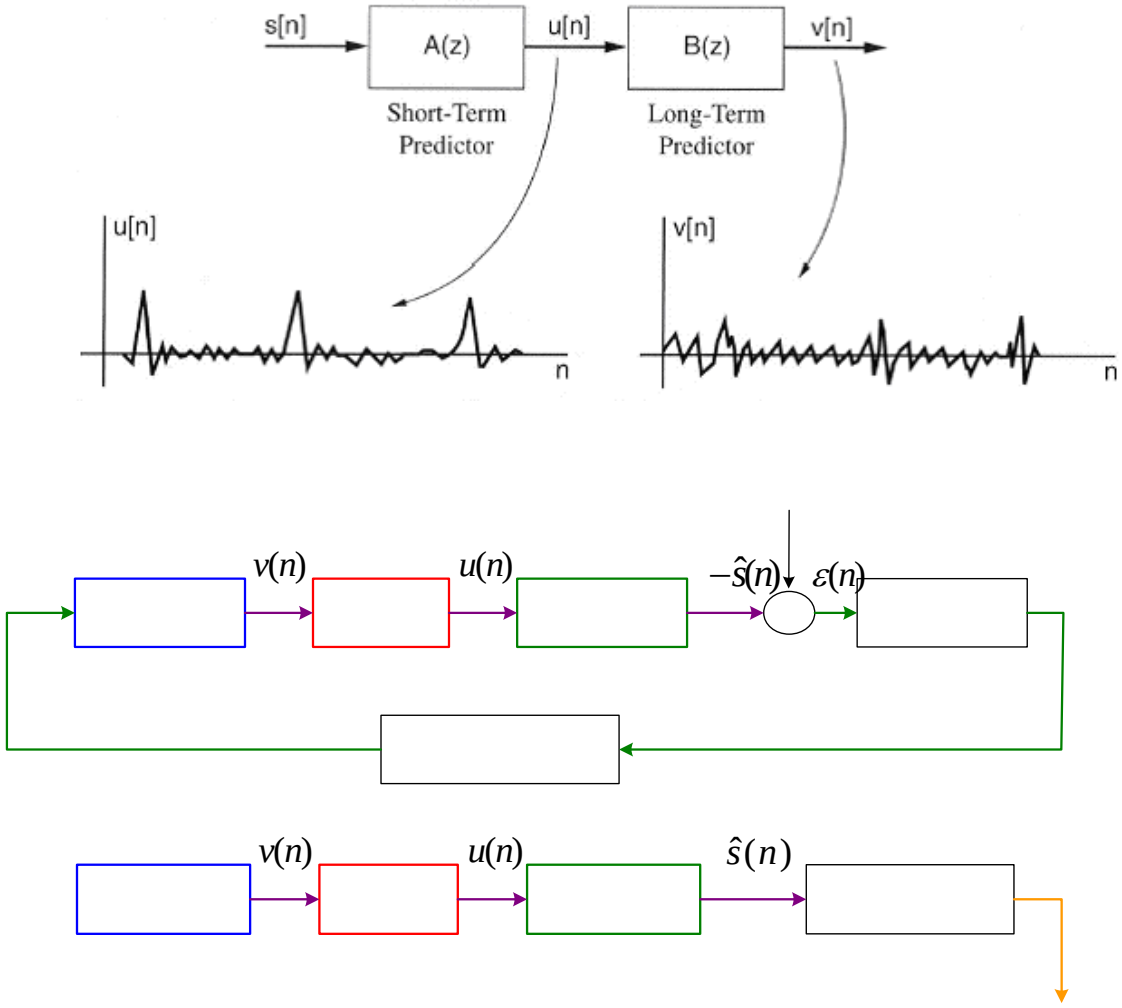


Illustration of short-term followed by long-term prediction in AbS techniques. The residual signal *u*[n] is the short-term prediction error and the residual *v*[n] is the long-term prediction residual.



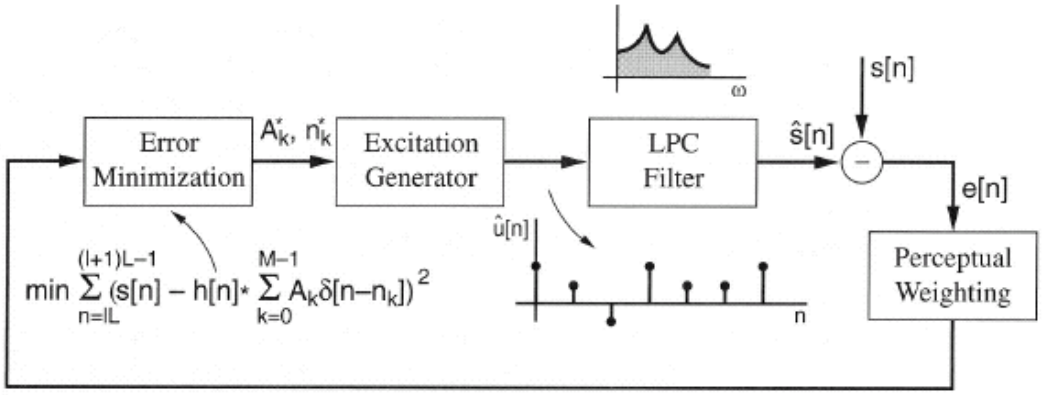
1. Long-Term Predictor (LTP):

$$\frac{1}{P(Z)} = \frac{1}{1-P_l(Z)} = \frac{1}{1-\sum_{k=-m_1}^{m_2} G_k . Z^{-(\alpha+k)}} \qquad \begin{cases} m_1 = m_2 = 0 & 1-tap\ predictor \\ m_1 = m_2 = 1 & 3-tap\ predictor \end{cases}$$

2. Short-Term Predictor (STP): Same as LPC-10 Federal-Standard 1015.
   - 10th Order LPC with Autocorrelation method following a Hamming windowing and a pre-emphasis.
   - LPC coefficients a_k are re-scaled by $a_i.\gamma^i$, where $\gamma = 0.994$, which shifts the poles of $1/A(Z)$ towards the origin.
   - Line-Spectral Pair (LPS) transformations are done on $a_i.\gamma^i$ for efficient coding.

Multi-pulse Excitation AbS Speech Coding

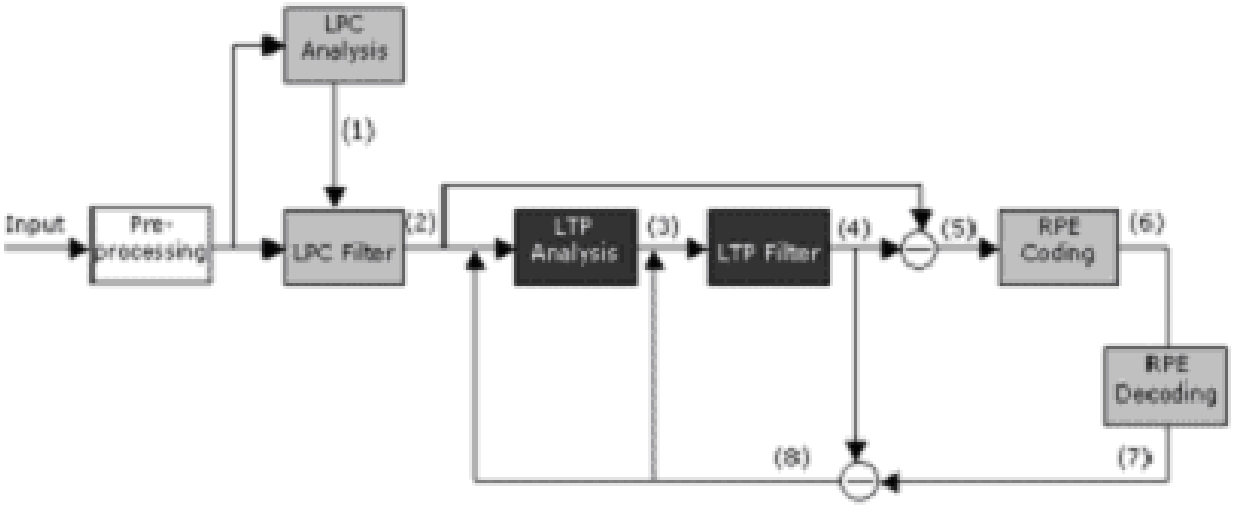
Instead of a single pulse excitation to represent the glottal pulse, several of them are employed to improve the sound produced by the analysis-by-synthesis codecs.



Multi-pulse linear prediction as an analysis-by-synthesis approach.

13 KB Group Special Mobile (GSM) Full-Rate Codec:

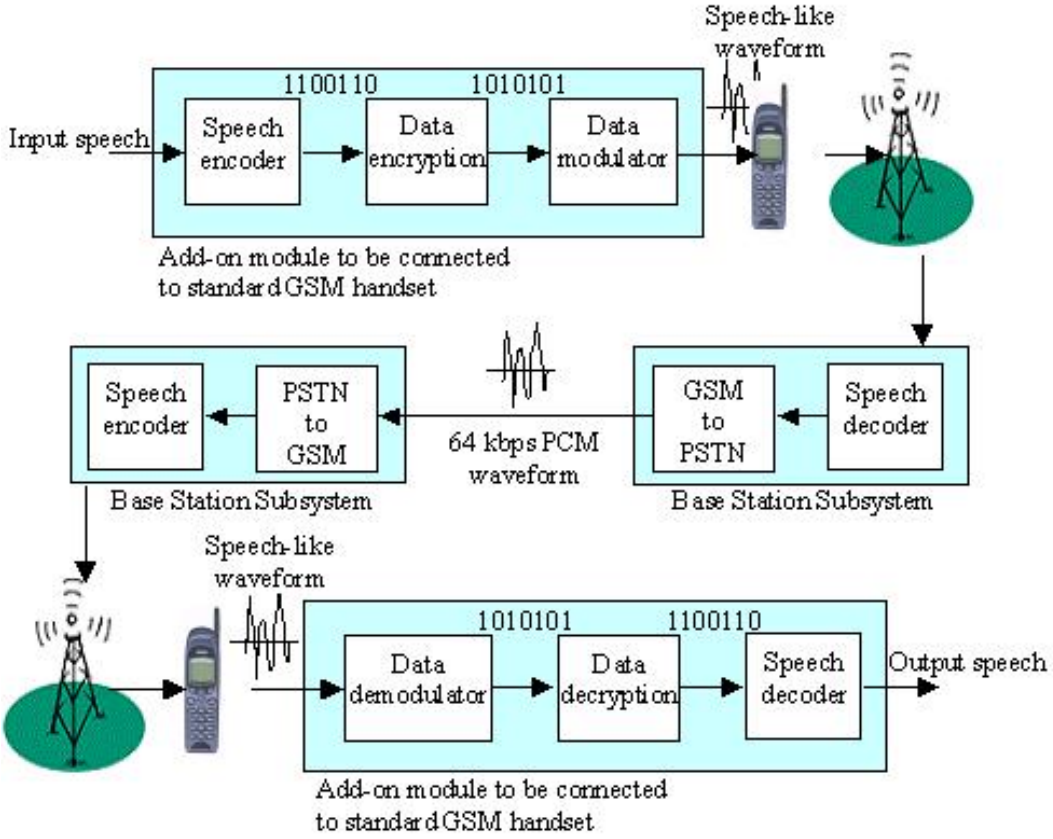
- Pan European Standard, which uses Regular Pulse Excitation (RPE) with LTP.
- Full-Rate Operates at 13.0 KB/s and the half-rate has been introduced recently at 6.5 KB/s.
- Quality: Mean Opinion Score (MOS) for 13 KB/s: 3.5-3.9 (Max: 5.0)
- Computational Complexity: 5-6 MIPS.



(1) Reflection coefficients are converted to log area ratios for transmission
(2) Short-Term residual signal
(3) LTP lag and gain parameters
(4) Short-term residual estimate (STP)
(5) Short-term residual signal
(6) RPE parameters
(7) Reconstructed long-term residual signal
(8) Reconstructed short-term residual signal.

Parameter	Number of parameters	Total bits per frame
LARs	8 per frame	36 bits
LTP lag	1 per subframe (7 bits)	28 bits
LTP gain	1 per subframe (2 bits)	8 bits
RPE grid position	1 per subframe (2 bits)	8 bits
Block amplitude	1 per subframe (6 bits)	24 bits
RPE Pulses	13 per subframe (3 bits each)	156 bits
Total		**260 bits per frame**

GSM Codec Implementation in Cellular Networks:

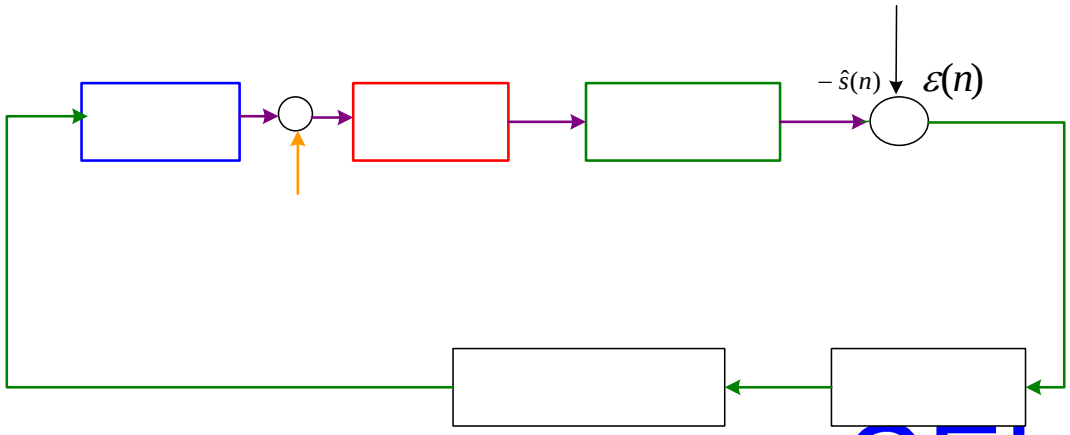


Code-Excited Linear Prediction (CELP) Codecs

The principle is similar to the Analysis-by-Synthesis LPC Codec *except*: vector quantization (VQ) based code excitation is used. The codebook has 512-1024 codewords with K=60 dimensions, randomly generated.

- Frame size is 30 msec (240 samples)
- Excitation $u(n)$ is coded directly
- Higher rate and quality system
- Computationally more complex
- Pitch prediction filter has replaced the LTP.

Block diagram of the CELP encoder:



CELP En

- The pitch prediction (LTP) filter is given by:

$$\tilde{H}(Z) = \frac{1}{1 + bZ^{-T}}$$

where ***T*** could be an integer or a fraction thereof.

- The perceptual weighting filter is given by:

$$W(Z) = \frac{H(Z/\gamma_2)}{H(Z/\gamma_1)}$$

where $\gamma_1 = 0.9; \gamma_2 = 0.5$ have been determined to be good choices.

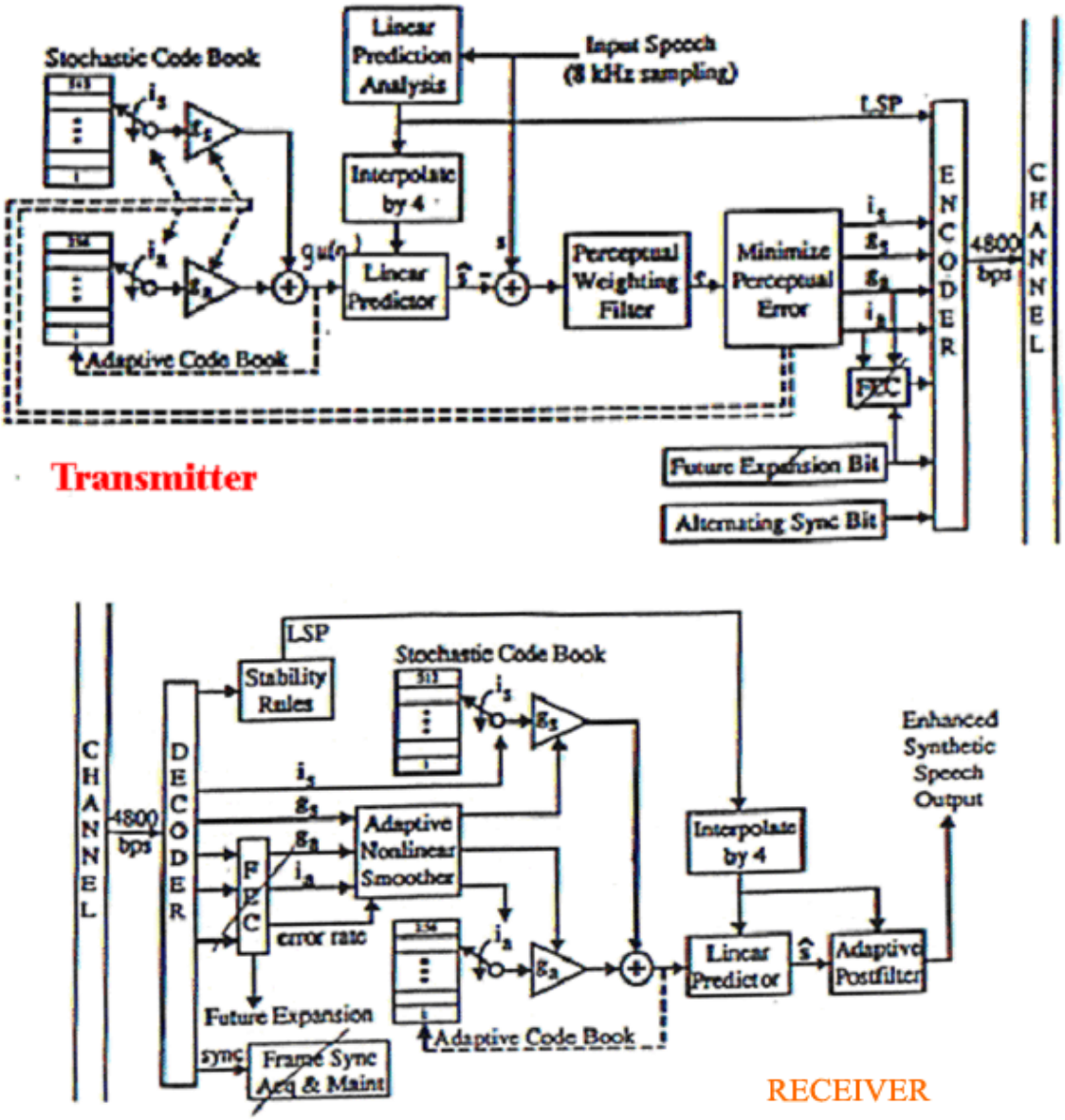
- Each frame is divided into 4 sub-frames for pitch (LTP analysis) and excitation. In each sub-frame, the codebook contains 512 codevectors.
- The gain is quantized using 5 bits per sub-frame.
- The LSP parameters are quantized using 34 bits similar to the LPC Codec of Fed STD 1015..
- At 30 ms/frame, 4.8 kbps is equivalent to 144 bits/frame. 144 bits are allocated as follows:

Random Codebook

Parameter	No. of Bits
LSP	34
Pitch Prediction Filter	48
Codebook Indices	36
Gains	20
Synchronization	1
FEC	4
Future Expansion	1
Total	**144**

Choose next s

Fed-STD:1016 Compatible CELP Codec (Stochastic & Adaptive Codebooks)



1. Adaptive Codebook:
   - 256 vectors of 60-samples long (60-Dimensional VQ Vector)
   - Codebook search is performed by closed-loop analysis-by-synthesis.
   - 256 code vectors correspond to 128 delays and 128 non-integer delays, ranging from 20 to 147 samples.
   - Non-integer delay code vectors can be obtained by interpolation of integer-delay code vectors.
2. Stochastic (Random) Codebook:
   - 512 vectors of 60-samples long.
   - Code vectors are sparse, overlapping, ternary-valued and pseudo-randomly generated by 2-shift and overalp technique.

3. Short-Term Prediction (STP)
   - 10^{th} Order Autocorrelation analysis using Hamming Window.
   - LPC coefficients a_k are re-scaled by $a_i.\gamma^i$, where $\gamma = 0.994$, which shifts the poles of $1/A(z)$ towards the origin.
   - 10 LSP coefficients are obtained from LPC as in earlier systems.
4. Complexity: 16 MIPS.
5. Quality: MOS for 4.8 KB/s: 3.2
6. Used by Department of Defense (DoD) in the US and NATO in their 3rd Generation secure telephone units (STU-III).

	Linear Predictor	Adaptive CB	Stochastic CB
Update	30 ms	30/4 = 7.5 ms	30/4 = 7.5 ms
Parameters	10 LSPs (independent)	1 gain, 1 delay 256 codewords	1 gain, 1 index 512 codewords
Analysis	open loop 10th order autocorrelation 30 ms Ham window no preemphasis 15 Hz expansion interpolated by 4	closed loop 60 dimensional mod MSPE VQ weighting = 0.8 delta search range: 20 to 147 noninteger delays	closed loop 60 dimensional MSPE VQ weighting = 0.8 shift by -2 77% sparsity ternary samples
Bits Per Frame	34 (3,4,4,4,4,3,3,3,3,3)	index: 8+6+8+6 ±gain: 5 x 4	index: 9 x 4 ±gain: 5 x 4
Rate	1133.33 bps	1600 bps	1866.67 bps
Miscellaneous	The remaining 200 bps are used as follows: 1 bit per frame for synchronisation, 4 bits per frame for forward error correction, and 1 bit per frame for future expansion.		

4.8 KB/s CELP Demo

8.0 KB/S CS-ACELP ITU-G729 and G.729A Standards

- CS-ACELP=Conjugate-Structured Algebraic CELP.
- The principle is similar to the 4.8 kbps CELP Coder *except*:
  - Frame size is 10 msec (80 samples)
  - There are only two subframes, each of which is 5 msec (40 samples)
  - The LSP parameters are encoded using two-stage vector quantization.
  - The gains are also encoded using vector quantization.
- At 10 msec per frame, 8 kbps is equivalent to 80 bits/frame. These 80 bits are allocated as follows:

Parameters	**No. of Bits**
LSP	18
Pitch Prediction Filter	14
Codebook Indices	34
Gains	14
Total	**80**

8.0 KB/S CS-CELP Demo.

8.0 Vector Sum Excited LP (VSELP) Standard

- Very similar to Fed-Std 1016 operating at 8,000 bits/s.
- Embedded to 3rd generation cellular standards under the code:
- Highly structured codebooks to reduce complexity and to increase robustness to channel errors.
- Quality: MOS at 6.3 and 8 KB/s: 3.4 and 3.5, respectively.
- Complexity: 14 MIPS.

IS-96 Industry Standard used in CDMA cellular phones.

- CELP with STP
- Variable bit rate operating at 1.2, 2.4, 4.8, and 9.6 KB/s.
- Quality for 9.6 KB/S: 3.3
- Complexity: 15 MIPS

ITU G.728 Low-Delay CELP 16 KB/s Standard

- Very similar to Fed-Std 1016 uses CELP with a low-delay constraint.
- Shorter frames and shorter excitation vectors.
- Quality: MOS at 16 KB/s: 3.4
- Complexity: 30 MIPS

16.KB/S CS-CELP Demo and **64 KB/S Original Speech**

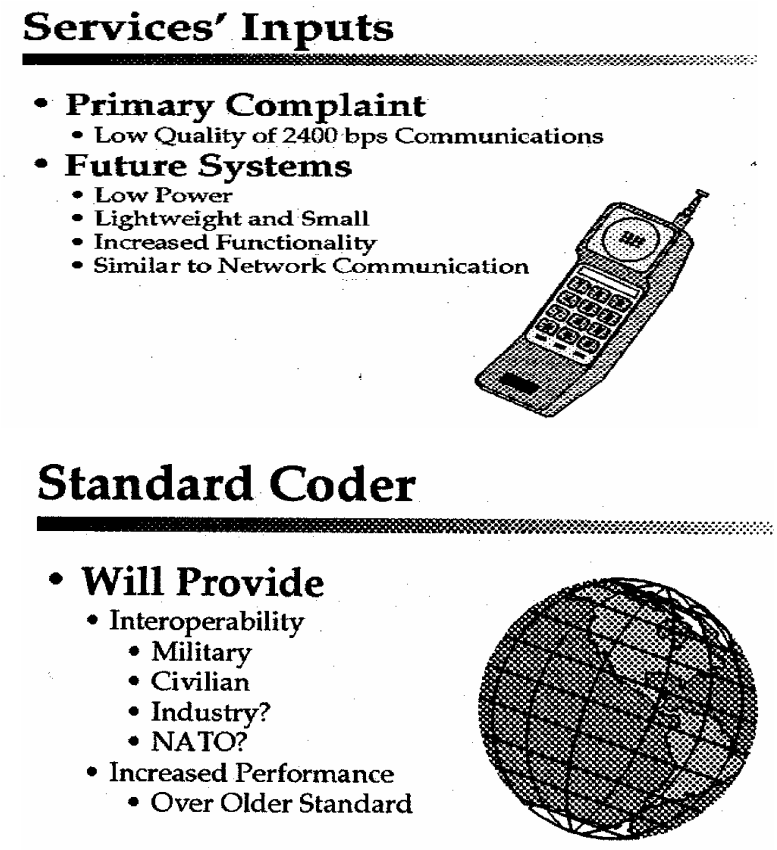
SUMMARY and MOS Performance:

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

MELP CODER

2400 bps US Government Speech Coding Standard:

- The quality of the synthesized speech in the 2400 bps Speech coder (FS1015 LPC-10, which is commonly used in military communication systems and in many commercial low-end speech codecs, has been very limited.
- FS1015 LPC-10 performs poorly in the presence of severe noise and/or bit error rates. (Cockpit and battlefield conditions and BER of 1-5% range).
- CELP FS-1016 vocoder (4.8 Kbps) also uses a 10th order LPC analysis for a "short term prediction" and an analysis-by-synthesis (AbS) architecture. Owing to the exhaustive codebook search and filtering needed to maximize the match score, the computational load demand is very high and many alternative codebook structures and search methods have been explored.
- It needed additional channel coding for error protection and encryption purposes, which resulted in additional bandwidth and power.
- Due to increased bandwidth requirement and overall system complexity and power requirement and it could not be used in low-bandwidth channels of battlefield, naval, and cockpit applications.
- A new 2400 coder became necessary.



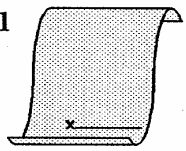
US Department of Defense has invited a number of research teams to come up with a 2400 bps speech coder to be the next Standard Codec. Initially (7) candidates competed. The requirements for the systems to be considered as the US Military Standard together with NATO countries were:

Equivalent or better than CELP on:

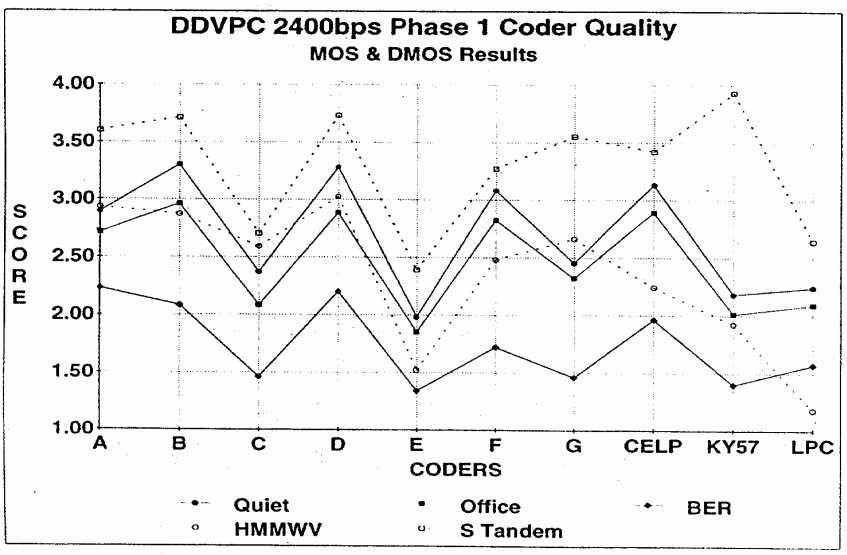
- Quality
- Intelligibility
- Robustness against environmental changes
- Complexity one DSP-Chip
- End-to-end maximum delay not to exceed 180 ms.

System Requirements

- **Match FS1016 4800 bps CELP performance**
- **Low power**
- **Tandem communications**
- **Talker and Environmental Independence**
- **Talker Recognizability**

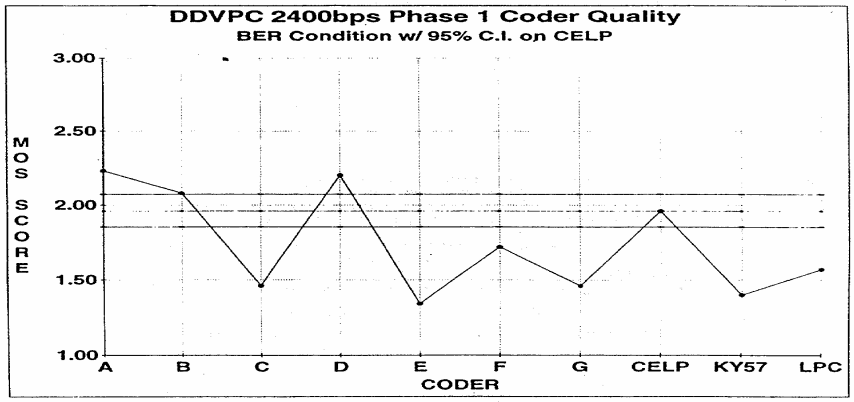


- Codecs were evaluated using a number of tests including MOS, DMOS, DRT, DAM, intelligibility, speaker ID, 1% random BER, complexity, inter-operability, and number of others.
- The benchmarks were CELP, LPC-10, US Air Force KY57 CVSD.
- MOS Results:



Vocoder Intelligibility & Quality Test Methods - Tardelli & Kreamer

- Performance in the presence of BER:



- Only FOUR candidates were invited to Phase 2 for full evaluation:
- Overall Results:

1996 DDVPC 2400bps Coder Evaluation
Figure of Merit (FOM) Calculations

Coder	A	B	C	D	CELP	LPC	Max
Performance	82.942	83.590	83.378	83.718	80.125	69.313	86.151
Improvement over Celp	2.817	3.465	3.253	3.593	0.000	-10.812	6.026
85% Perf. Improvement	2.395	2.945	2.765	3.054	0.000	-9.190	5.122
Complexity	82.488	84.013	81.022	80.448	80.028	79.607	85.952
Improvement over Celp	-2.460	-3.985	-0.994	-0.420	0.000	0.421	-5.924
15% Comp. Improvement	-0.369	-0.598	-0.149	-0.063	0.000	0.063	-0.889

FOM = 0.85*(Pcoder-Pcelp) + 0.15(C'celp-C'coder)

Figure of Merit	2.026	2.347	2.616	2.991	0.000	-9.127	4.234
Rank	4	3	2	1	5	6	
Requirements	Yes	Yes	Yes	No			
	A	B	C	D	CELP	LPC	Max
Coder ID	LL	ATT	TI	DVSI			
DDVPC RECOMMENDED STANDARD			TI				

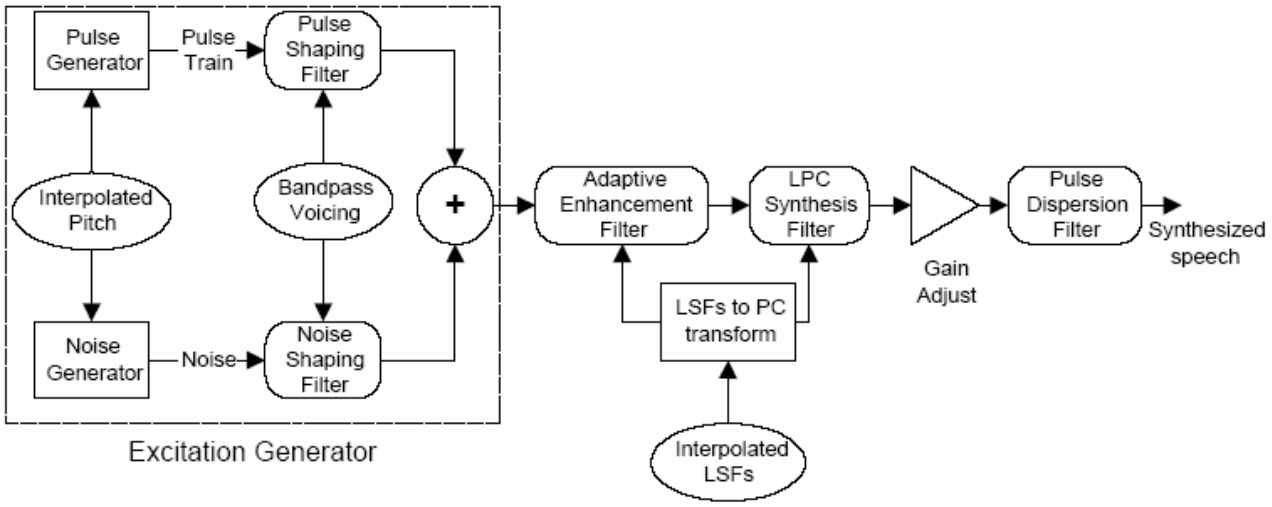
Conclusion: Although, Coder D from DVSI had the highest score overall but it failed one of the critical MUST conditions (non-negotiable). Thus, Coder C called ***MELP Coder*** from the consortium, *"Texas Instruments, Georgia Institute of Technology and Atlanta Signal Processing Inc."* has been chosen as the new standard.

MELP:

MELP encoder is generally similar to a classical linear prediction encoder. Main differences are in the synthesizer, which includes:

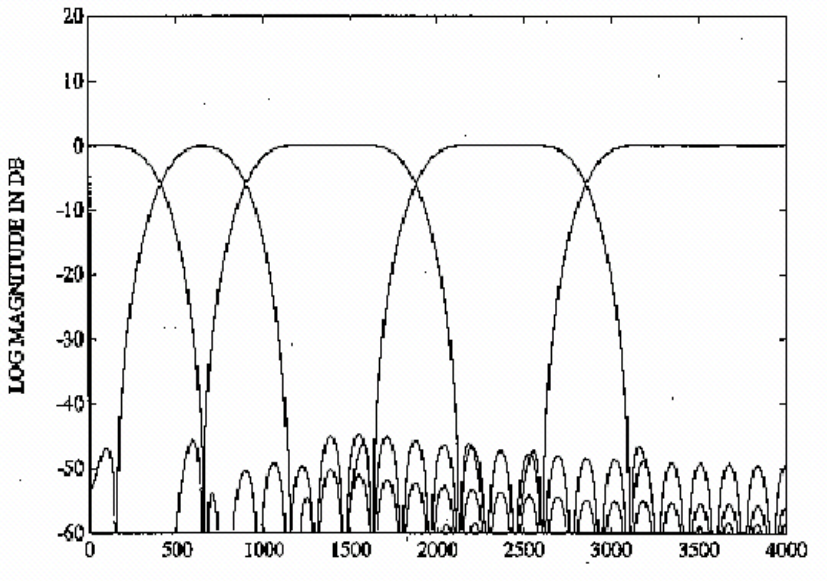
1. Conversion of linear prediction coefficients to line spectral frequencies for quantization.
2. A third voicing state for jittery voiced frames. No longer V/UV 2-state binary decision on voicing. Instead five band-pass voicings for noise/pulse mixture control at the synthesizer.
3. Synthesizer of the MELP decoder also includes adaptive enhancement filtering and a pulse dispersion filter that improves the match between synthesized and original voiced speech.
4. The “buzziness” which is a common problem in classical LPC speech is removed by the noise mixture algorithm, and the thumps, metallic and tonal noises due to voicing errors are removed using the third voicing state.

MELP SYNTHESIZER:



Key Features of Mixed Excitation:

1. Different mixes of pulse and noise in each of (4-10) frequency bands.
2. Pulse train and noise sequence is each passed through time-varying spectral shaping filters.
3. They are added together to give a full-band excitation similar to the glottal pulse in natural speech.
4. Aperiodic Pulses: Although, mixed excitation can remove the buzziness from synthesized speech, it does not eliminate *"short isolated tones"*. This is eliminated by adding *"noise"* in the lower frequency bands, which acts as a third *"voicing state."*
5. Adaptive Spectral Enhancement: It helps the bandpass filtered synthetic speech to match natural speech waveforms in the formant regions.
6. Pulse Dispersion Filter: It improves the match in frequency bands, which do not contain a formant resonance.



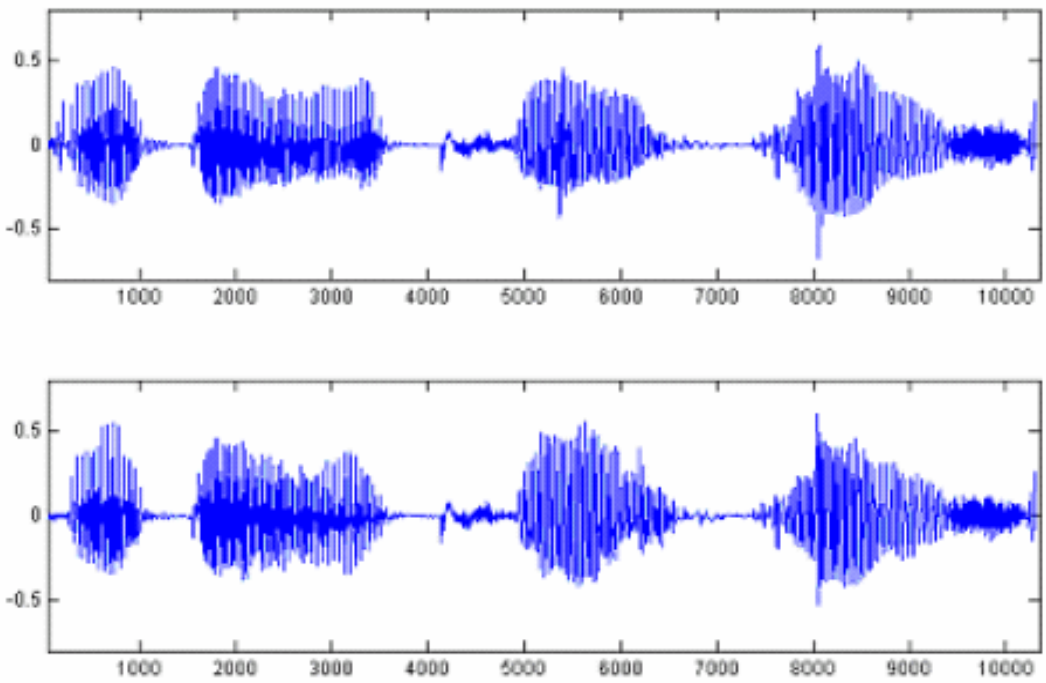
Example bandpass filter responses, 5 band, 48th order.

2.4 KB MELP Encoder:

1. Sampling Rate = 8000 samples/s.
2. Frame length = 22.5 ms.
3. LPC order = 10; Autocorrelation model.
4. LSP parameters coded by scalar quantizers.
5. Pitch estimated from a search of normalized correlation coefficients of the LP filtered residual signal of BW=1200 Hz and explicit check for pitch doubling was performed.
6. Gross pitch errors were corrected using one past and one future frame information.
7. Overall voicing was decided on the basis of pitch periodicity.
8. Gain is estimated twice per frame using RMS value of the input speech.
9. 54 bits per frame at a rate 1/22.5ms=2,400 bits/s.

Table 1: 2.4Kbps MELP Parameter breakdown	
PARAMETER:	**NUMBER OF BITS/FRAME**
LSF's	**25 Bit Multistage VQ**
10 Fourier Magnitudes	**8 Bits VQ**
Pitch	**7 Bits**
Bandpass Voicing	**4 Bits**
Gain	**2 x 4 Bits**
Aperiodic Pulse	**1 Bit**
Sync	**1 Bit**

10. Original and synthetic speech waveform segments:



11. Diagnostic Acceptability Measure (DAM) score are significantly over LPC-10 (47.0) and it is comparable with the CELP at 4,800 bps.

Speech Coder	6 Speaker	Male	Female
2400 bps DoD LPC-10e	54.0	54.7	53.2
2400 bps ME LPC	58.9	59.8	57.7
4800 bps DoD CELP	62.6	63.5	61.5
4800 bps ME LPC	61.6	61.6	61.3

Table 2: Clean input DAM test scores

12. Extensions on both higher rates 4.8, 6.4 and 8.0 KB/s and lower rates 1.7 and 1.2 KB/s were developed as part of the annexes of the standard.
13. A single DSP chip implementation is extensively described in a report by Tan and Teo: Real-Time implementation of MELP Vocoder, J. of IE, Singapore, Vol. 44, No: 3, 2004.
14. 2400 KB/s Speech Demos (male, female)