Hybrid Speech Coding

Nimrod Peleg
Update: Jan 2004

Hybrid Coders

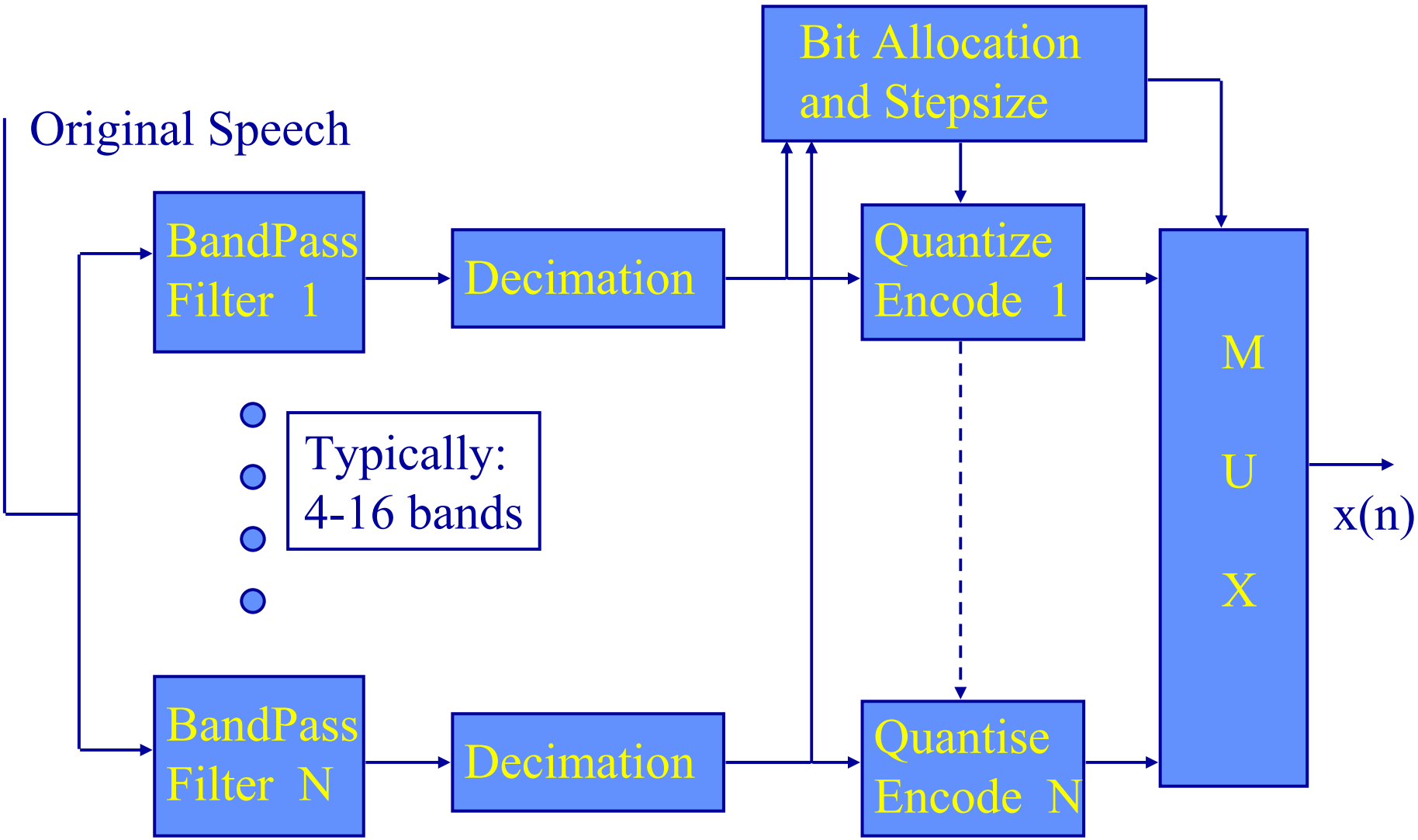
- Incorporate advantages offered by the basic schemes (Vocoders and Waveform coders)
- Two major categories:
  - Frequency domain: divide speech spectrum into bands or freq. components (SBC, ATC, MBE)
  - Time domain: Usually a linear prediction is involved, and speech is modeled by source-filter model (APC, RELP, MP-LPC, CELP, SELP, MELP)

Frequency Domain Coders

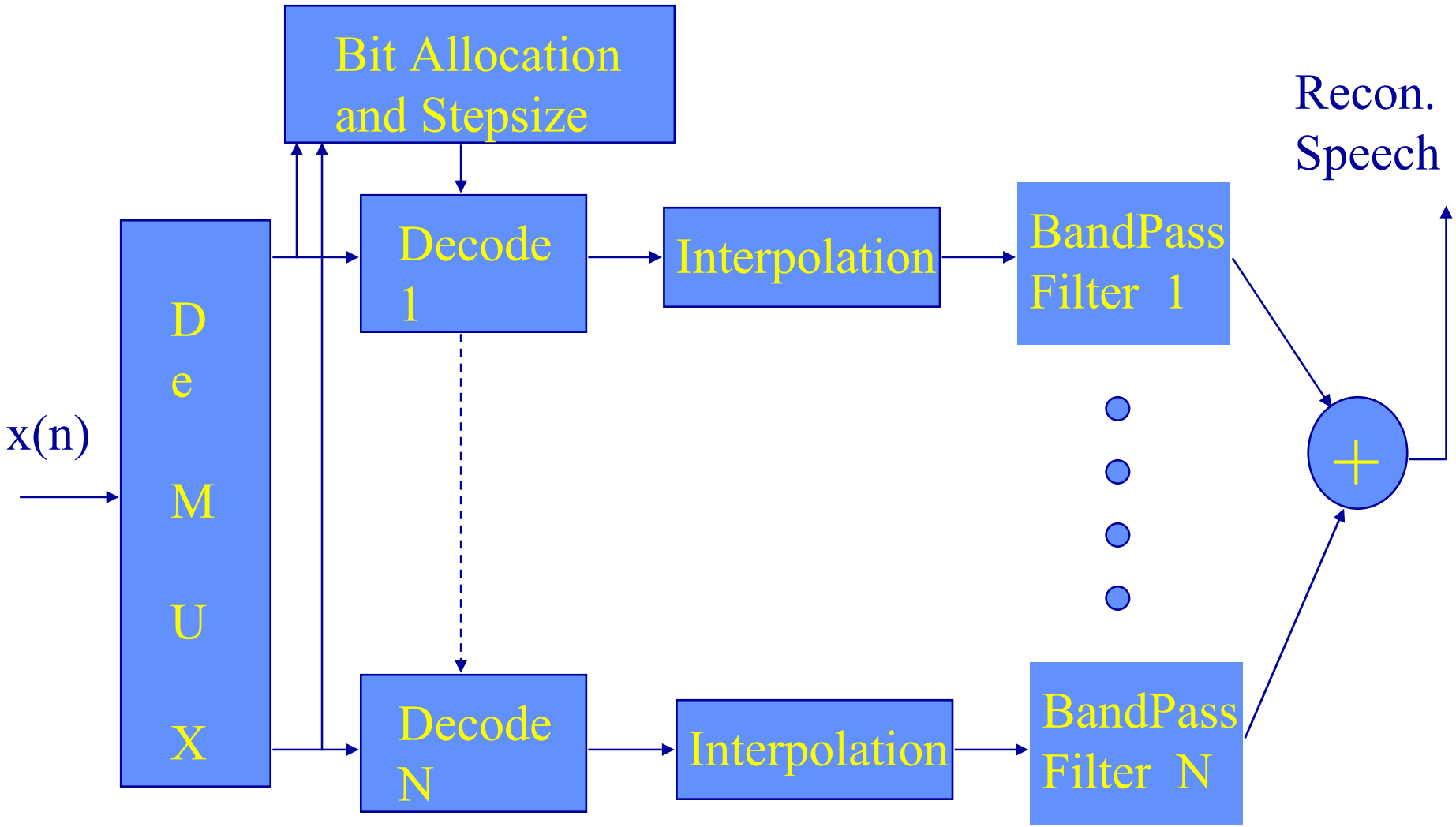
- After encoding and decoding, a replica of the input speech is synthesized (using the frequency components) by either filter bank summation or inverse transform.

- Basic assumption: signal varies slowly, so it can be modeled by short-time spectrum

Sub-Band Coding: Encoder



Sub-Band Coding: Decoder



Sub-Band Coding Features

- Short-time formant of speech spectrum can be exploited
- Quantization levels adopted to specific band
- Noise shaping can be applied
- Dynamic bit allocation
- Absence of pitch predictor causes rapid reduction in quality below 9.6Kbps

Bit Allocation

- Fixed allocation (early designs): number of bits for each sub-band determined from long-term signal statistics, and fixed for a given coder
- Adaptive Allocation (better results): number of bits assigned to each sub-band can vary according to local signal statistics (over 10-30mSec of speech segment)

Optimum Bit Assignment

Two major approaches:

- Minimum MSE criterion, based on signal variance, number of bands, noise energy etc.
- Allocation according to energies of each sub-band:
  - find band with largest energy
  - divide this energy by factor (~2) and allocate on bit
  - if not all bits allocated: repeat...

Noise Shaping

Noise is perceptually masked by a signal in the same frequency:

Use A Psychacoustic model

Adaptive Transform Coder (ATC)

- Process frames of about 20-30mSec
- Transform to freq. domain (DCT)
- Typical transform size: 128-256 samples
- Adaptive quantization to each coefficient, based on spectral envelope
- Bit allocation
- Better performance than SBC (transparent quality @16Kbps)
- Higher complexity and delay than SBC

Time Domain Coders

- Time domain coders vary in the way they treat the excitation signal of the time-varying filter
- Most of the bits are used for residual signal coding, so a family of *residual coding* schemes developed from the basic APC scheme

Adaptive Predictive Coder (APC)

- Employs both short and long-term linear prediction
- Excitation signal is inverse filtered and than scalar quantized (sample-by-sample)
- 16Kbps and below
- First: Attal and Schroeder, 1979

Residual Coding

Problems with LPC vocoders:

- Synthetic Quality
  - Spectral mismatch
  - Excitation model too simple (Buzz-Hiss)
  - Needs pitch detection and V/UV decision
- Non-Robust
  - Background noise (environment dependent)
  - deferent quality for mail / female
  - Multiple speaker problems

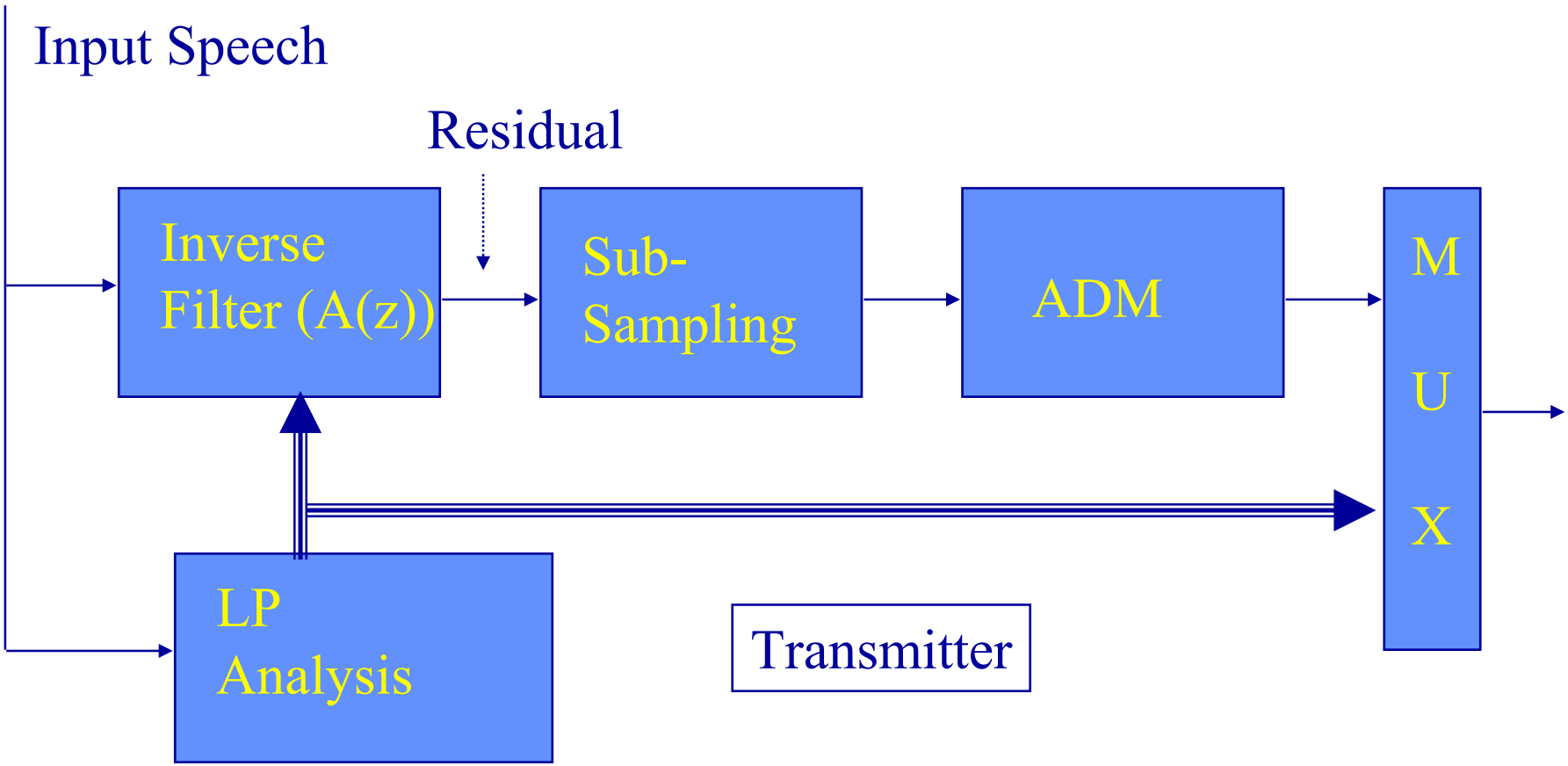
Residual Coding (Cont'd)

- Linear prediction filter acts as a short-time decorrelator: prediction residual has a flat spectrum, and is the perfect excitation for the synthesis filter
- The residual excitation signal carries all information not captured by LP analysis: phase, pitch, nasal sounds etc.

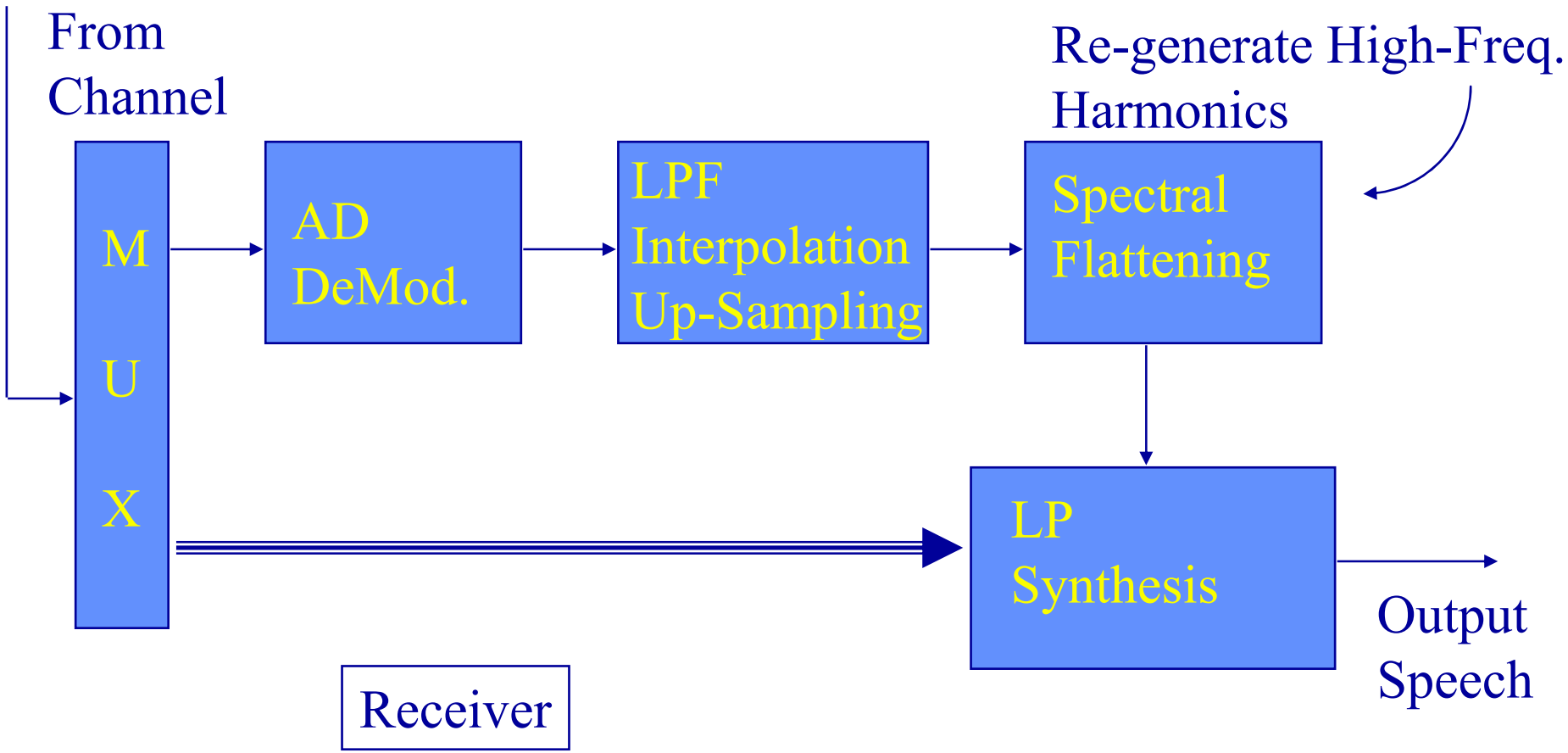
Residual Excited LP (RELP)

- RELP is different from ADPCM and APC (which also utilize prediction residual) since the residual encoding is is based on spectral matching (rather than waveform)
- RELP coders rely on the fact that low-freq. components are perceptually important
- For the following scheme: Residual bandwidth compressed to 800Hz (5Kbps)

RELP Encoder Scheme



RELP Decoder Scheme



RELP Limitation

- RELP quality is limited because information is lost in the residual baseband filtering