University of California
Berkeley

College of Engineering
Department of Electrical Engineering
and Computer Sciences

Professors : N.Morgan / B.Gold
EE225D

Spring,1999

Low Rate Coding

Lecture 25

Digital Vocoder

Date Rates	*Applications & Performance*	*Algorithms*
30,000 bps	Wide-band, High Fidelity Speech Transmission	Adoptive Differential Pulse Code Modelation [Standard?]
10,000-20,000 bps	Medium Band, Good Quality Speech Transmission. Some Noise Immunity	Split Band Systems (Part Modelling, Part Waveform Coding)
5,000-10,000 bps	Reasonable Quality Telephone Speech	Voice Excited Channel Vocoders VELP & RELP Adaptive Predictive Coding.
2,000-5,000 bps 2,400 bps (standard)	More Vulnerable to Environment. Speeker I.D.? Secrecy. Pitch Errors Restricted Bandwidth.	CELP, Multipulars, STC Channel Vocoders, LPC, Cepstral. Speech Dizitization Algorithms.
1,000-2,000 bps	Secrecy. Very Restricted Channels.	Frame-fill for Standard 2400bps Systems. Transformation to Reduce # of Bits for Parameters
500-1000 bps	Extremely Restricted or Busy Channels.	Vector Quantization
100-500 bps	Underwater Transmission? Super-Restricted Channels.	Recognition-Synthesis Phonetic Vocoder.

Frame - fill

* Start with a 2400 bps algorithm [Channel Vocoder].

* How do you get 2400bps?

Assume 400bps for excitation.

Assume 20ms frame or 50 frames/sec.

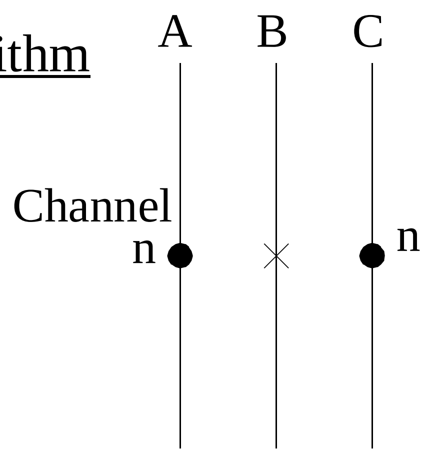
$$\frac{2000}{5\emptyset} = 40 \text{ bits/channel for a channel vocoder.}$$

Assume 16 channels $\frac{40}{16} = 2.5$ bits/channel, bits may be assigned on a perceptual basis. [Usually low channels need more bits.]

Question

How to reduce bit rate below 2400bps?

Algorithm



- Transmit A&C
- Reconstruct B at Receiver.

- If $B \approx A$, reconstruct $B = A$.
- If $B \approx C$, reconstruct $B = C$.
- If $B \approx A \notin B \approx C$, reconstruct B by averaging $A \notin C$.

Measurement are model at the transmitter.

Control bits tell Receiver how to handle the incomming data.

So 40bits gets reduced to 22. → Pitch can also be reduced.

$40 bits \times 50 = 2000$ $42 bits \times 25 = 1050$

Frame - fill for LPC

Whereas the channel vocoder synthesizer is always the same structure (a filler bank with malulaters for each b.p.filter) There are a few different LPC structures. For frame-fill, the trick is to find the best structure for finding a criterion of similarity.

a-parameters — direct form
k - parameters —lattice form
area ration
log area ration
→ ?

Experimental Result

Lattice structure works best.

DRT Results

Table IV

Pattern Matching on Vector Quantization

C.P. Smith (1960's) Buzo et al (1980's ?)

Consider the 2400bps channel vocoder.

Assume 40bits per frame.

* This corresponds to 2^{40} possible patterns.

Assume that 10^6 patterns are perceptually distinguishable.

If all 10^6 patterns were stored at both the receiver and transmitter.

Algorithm

- Compare incoming pattern with All 10^6 stared patterns.
- Send the 20 bit address of the best stared pattern.
- Reconstruct this best pattern at receiver.

Thus, bit rate is decreasing by 2:1.

If only 1000 patterns were perceptually distinguishable, we gain from 40 to 10 bits per frame. [4:1 Reduction]

Kany - Caulter 600bps Vocoder

Includes LPC analysis

clever LPC parameters to fewer formant parameters.

Vector quantization. - a form of frame fill.

Viewgraphs

Fig. 32.1 - Complete block diagram of Kamy - Coulter.

Fig. 32.2 - loci of poles is $K_{10} \rightarrow 1$

Fig. 32.3 - Spectrum of all pole system as poles migrates towards unit circle.

Vector Quantization - 128 formant patterns were stored.

(7bits) possible parameters to transmit.

Figure overall rate,
(including everything)
to get 600 bps.

Viewgraph of Lattice

Voiced
formants
PARCOR

Frame - Fill vs. Merging

In frame - fill, alternate frames are not sent - instead, Control bits are sent to direct the receiver to "optimally" recreate the missing frames. The sent frames can still be vector quantized.

In merging, the analyzer compares adjacent frames and decides whether these frames can be merged into a single frame. Perhaps more than just two adjacent frames can be merged. This means that the transmission rate is Variable, and Control bits are needed. Merged framed can also be vector quantized.

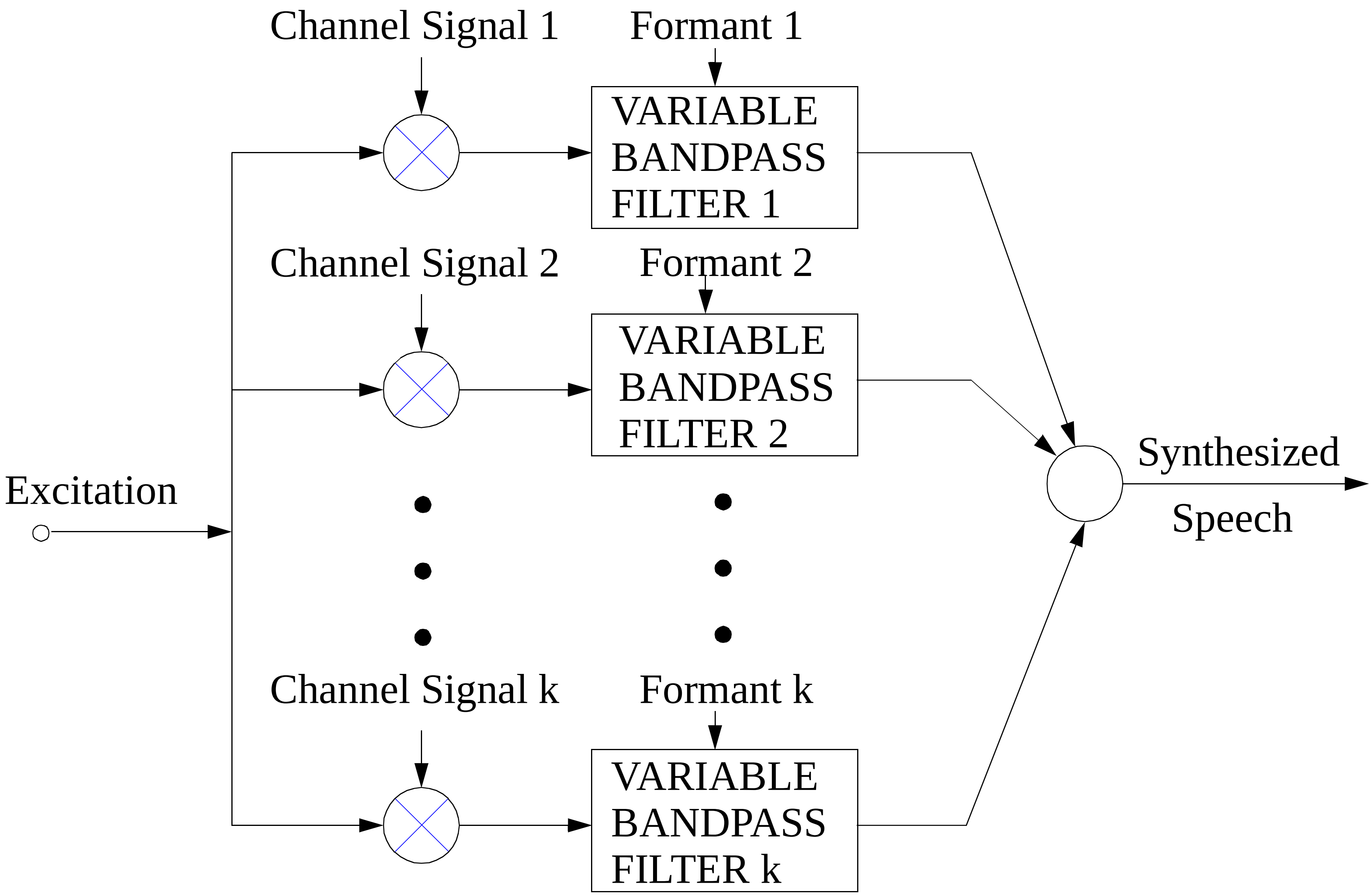


Figure 29.8: Parallel formant synthesizer.

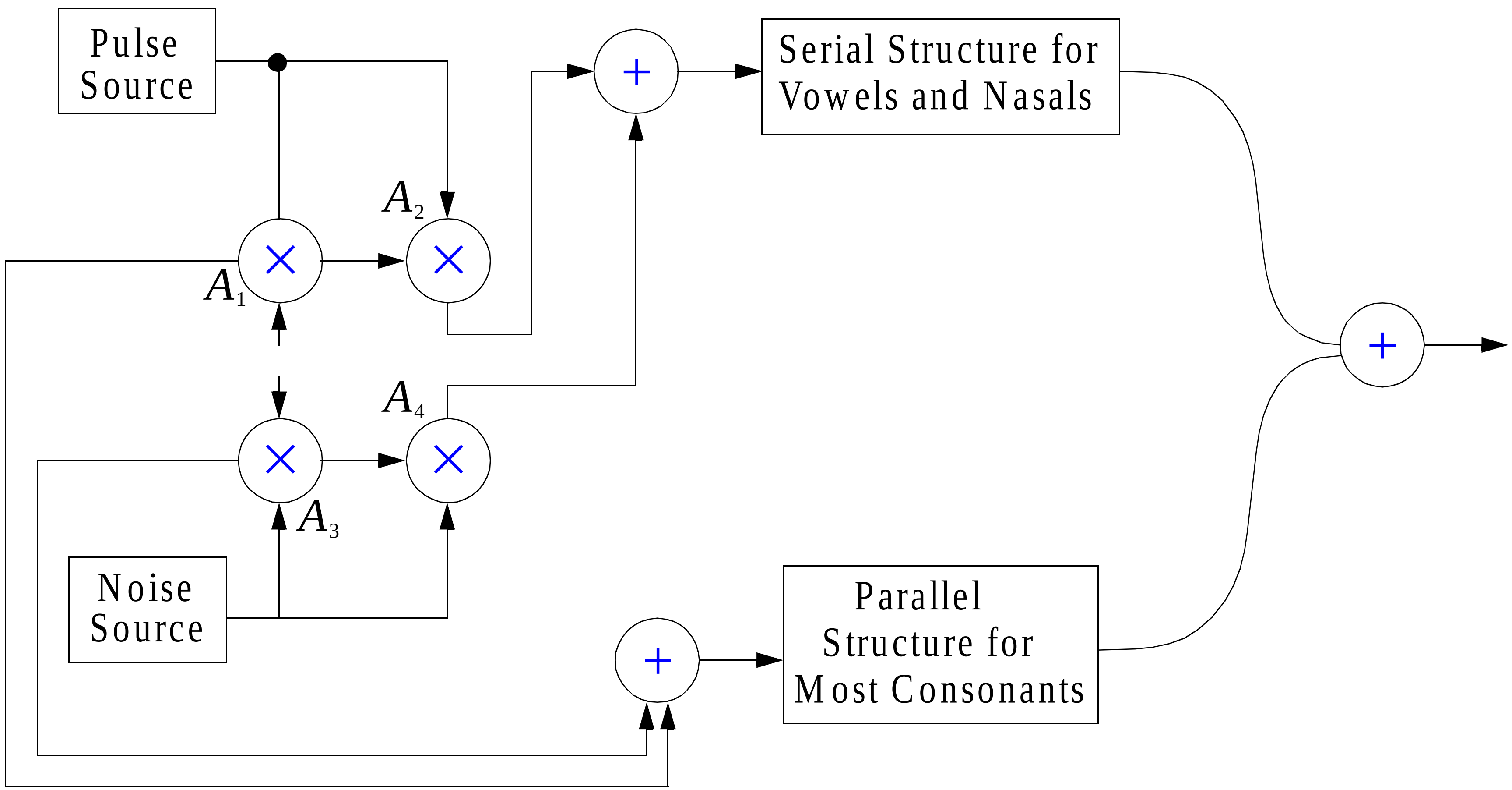


Figure 29.12 : Structure of Klatt Synthesizer.

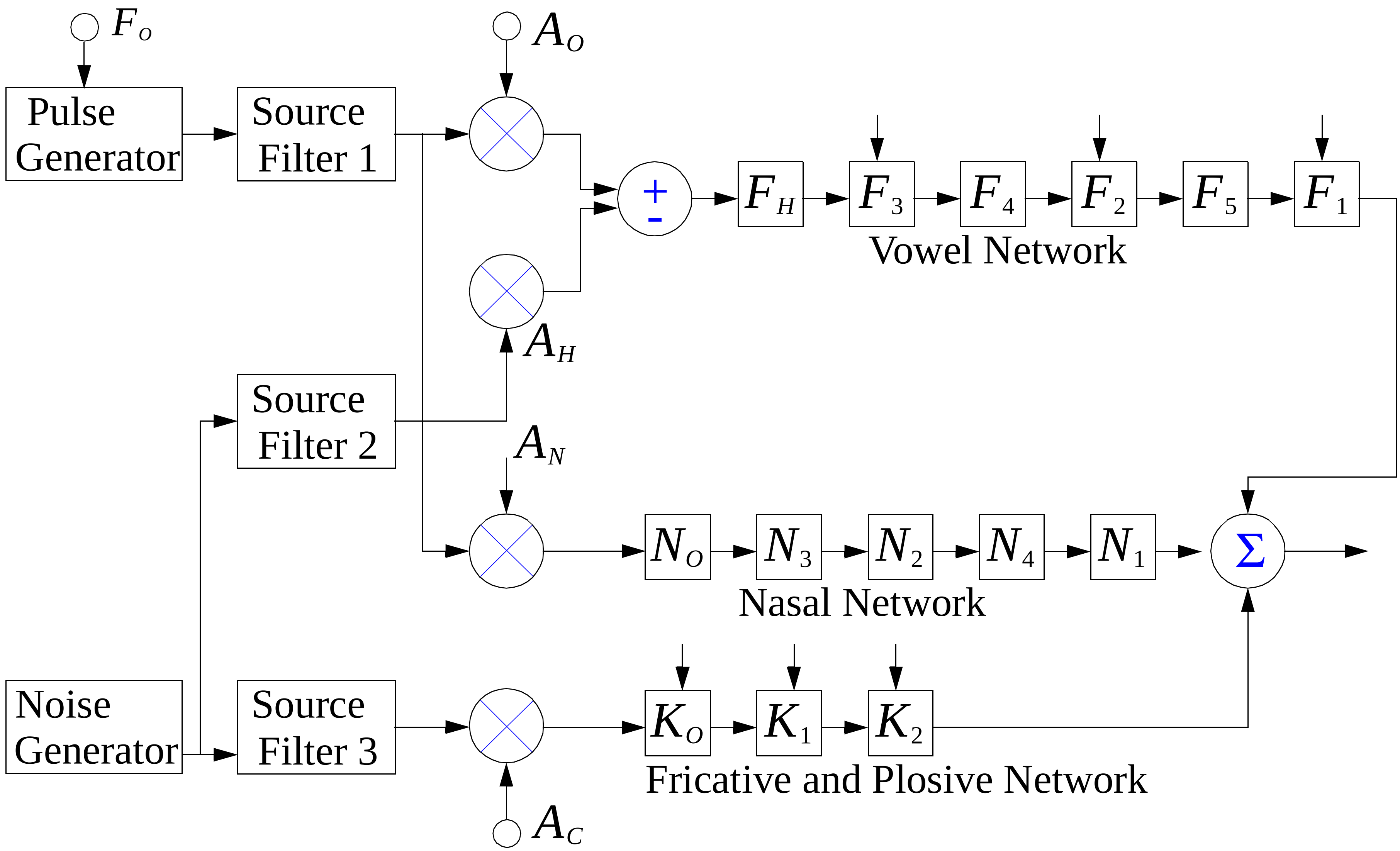
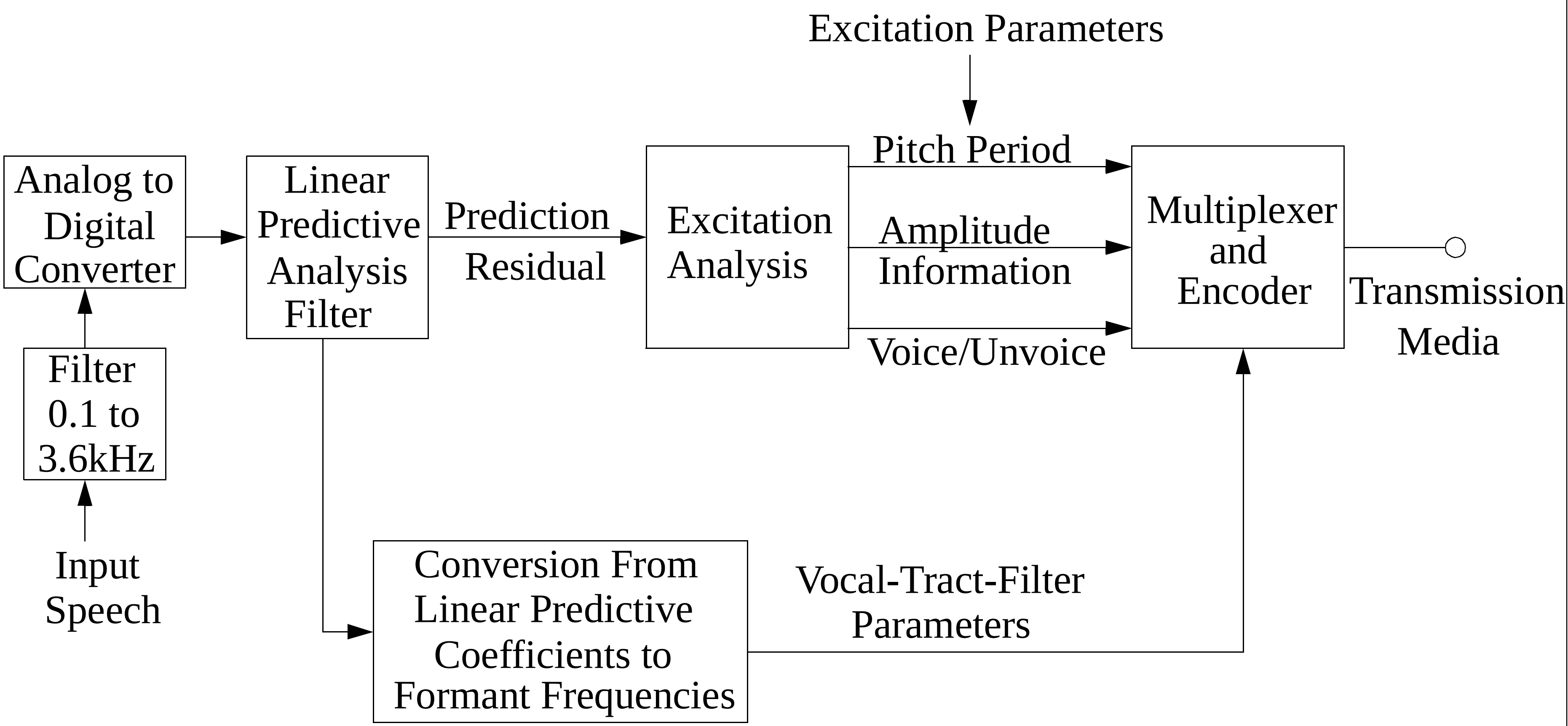
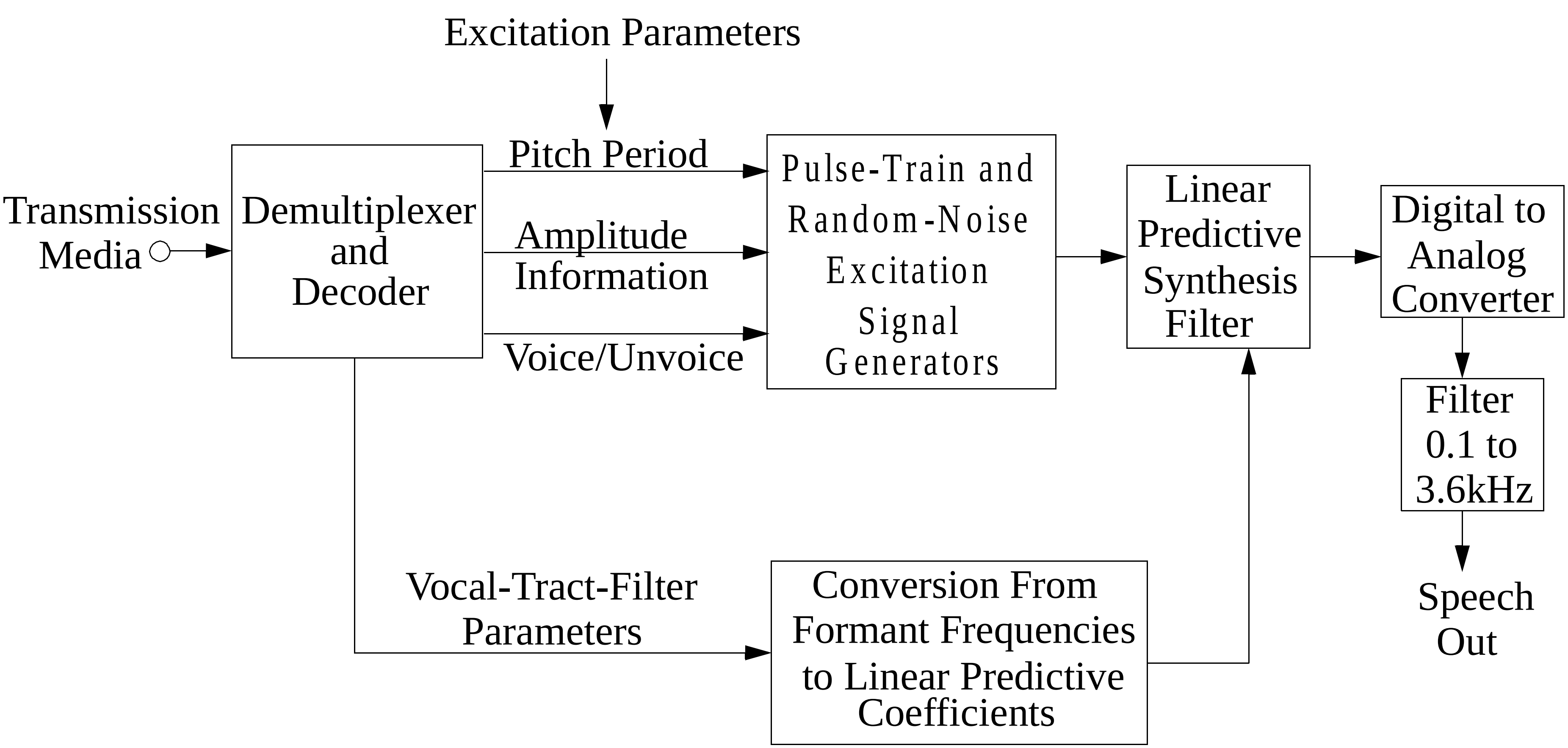


Figure 29.2: OVE II Speech Synthesizer of Gunnar Fant. Form [20]



(a) Transmitter

Figure 32.1 : Kang-Coulter 600 bps Voice Digitizer.



(b) Receiver

Figure 32.1 : Kang-Coulter 600bps Voice Digitizer

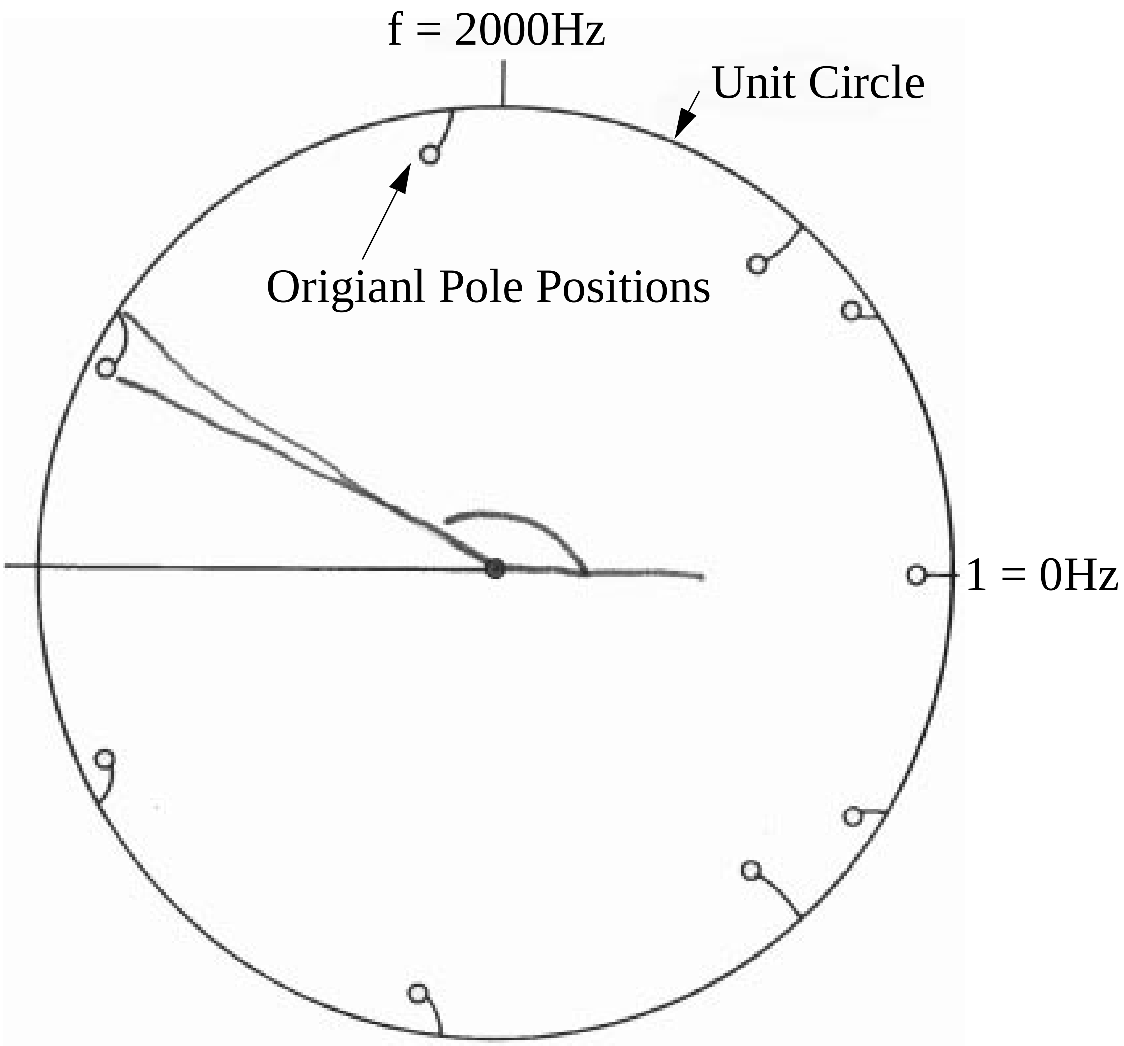


Figure 32.2 : Loci of the Poles as k_{10} Approaches Unity.

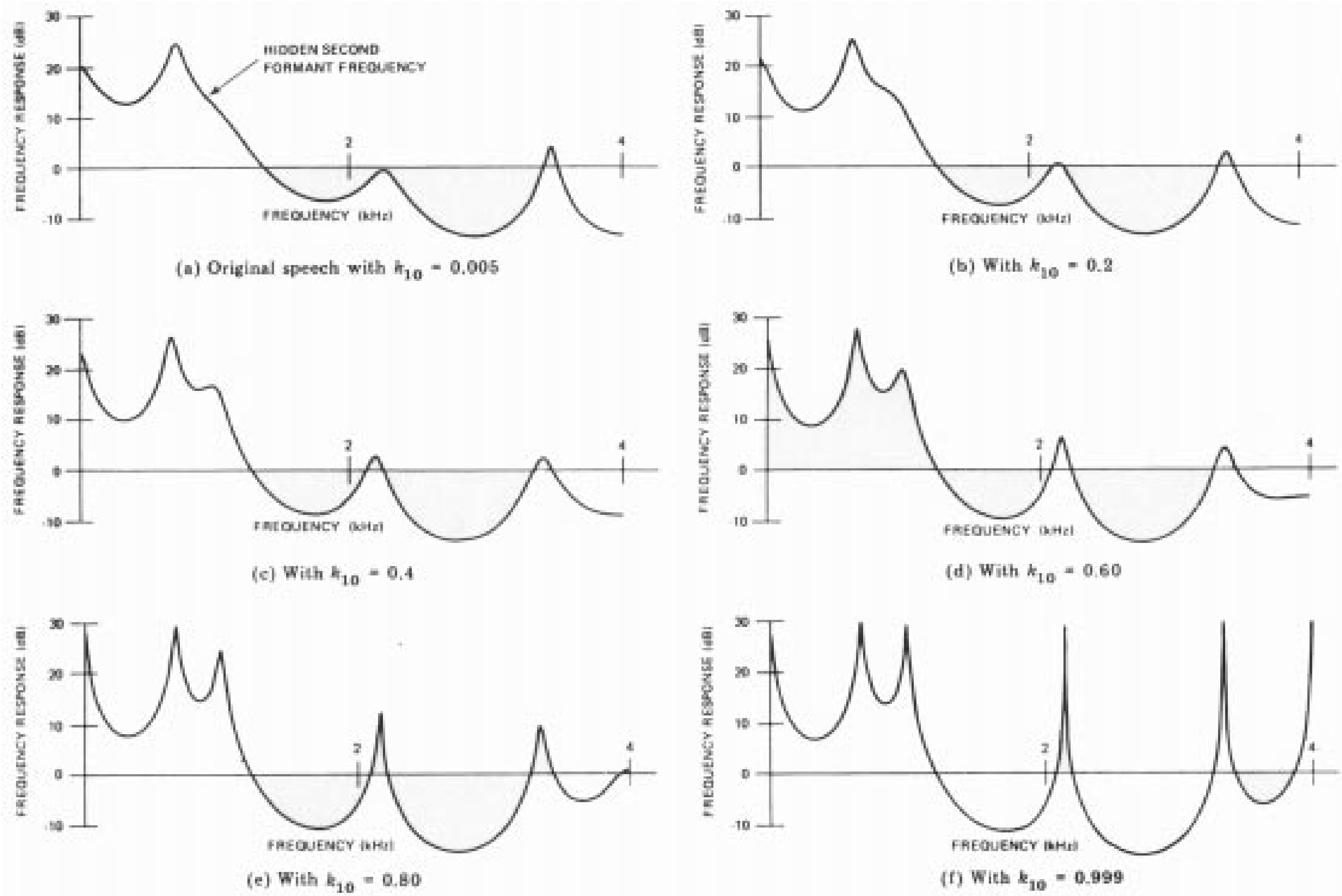
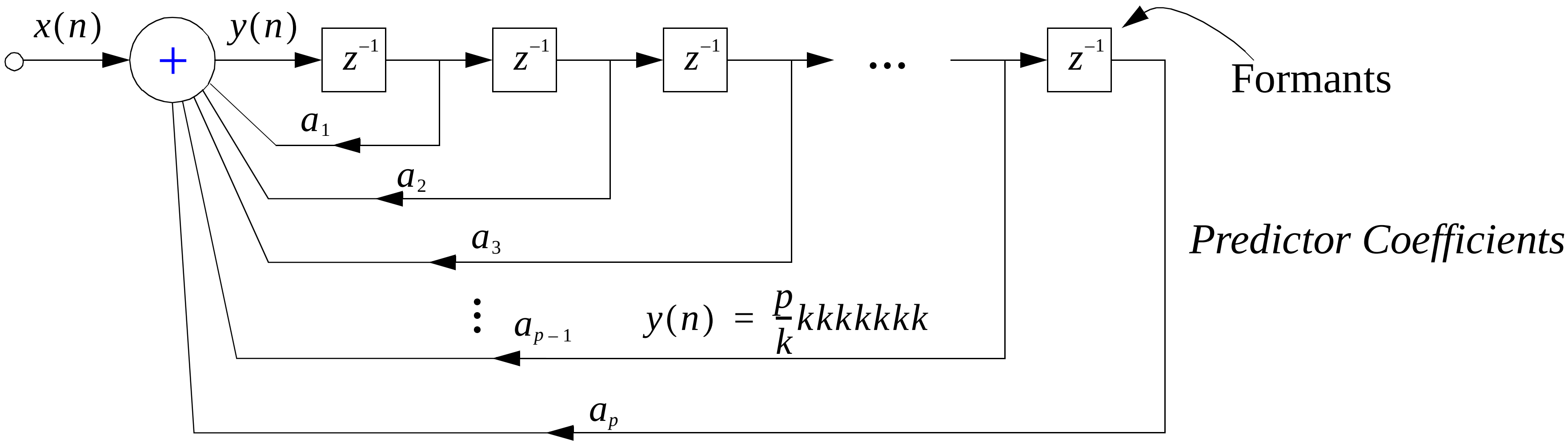


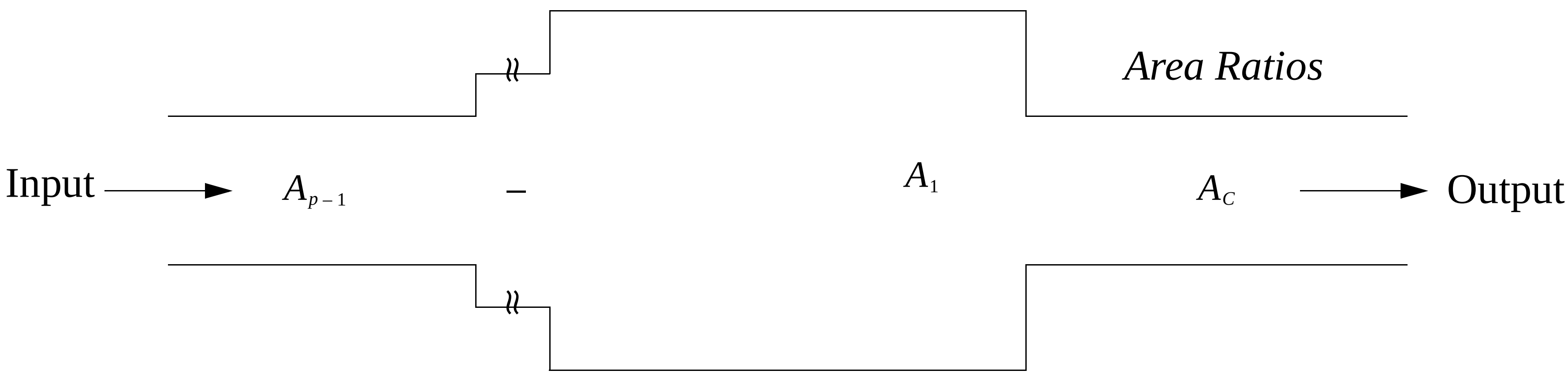
Figure 32.3 : Spectra of All-Pole System as Poles Migrate Towards Unit

$$Z^n - \alpha_1 Z^{n-1} - \alpha_2 Z^{n-2} \ldots - \alpha_n = \sum_{L=1}^{n} \Pi (Z - Z_i)$$

as $d_n \rightarrow 1$, Z_i' s migrate to unit circle.

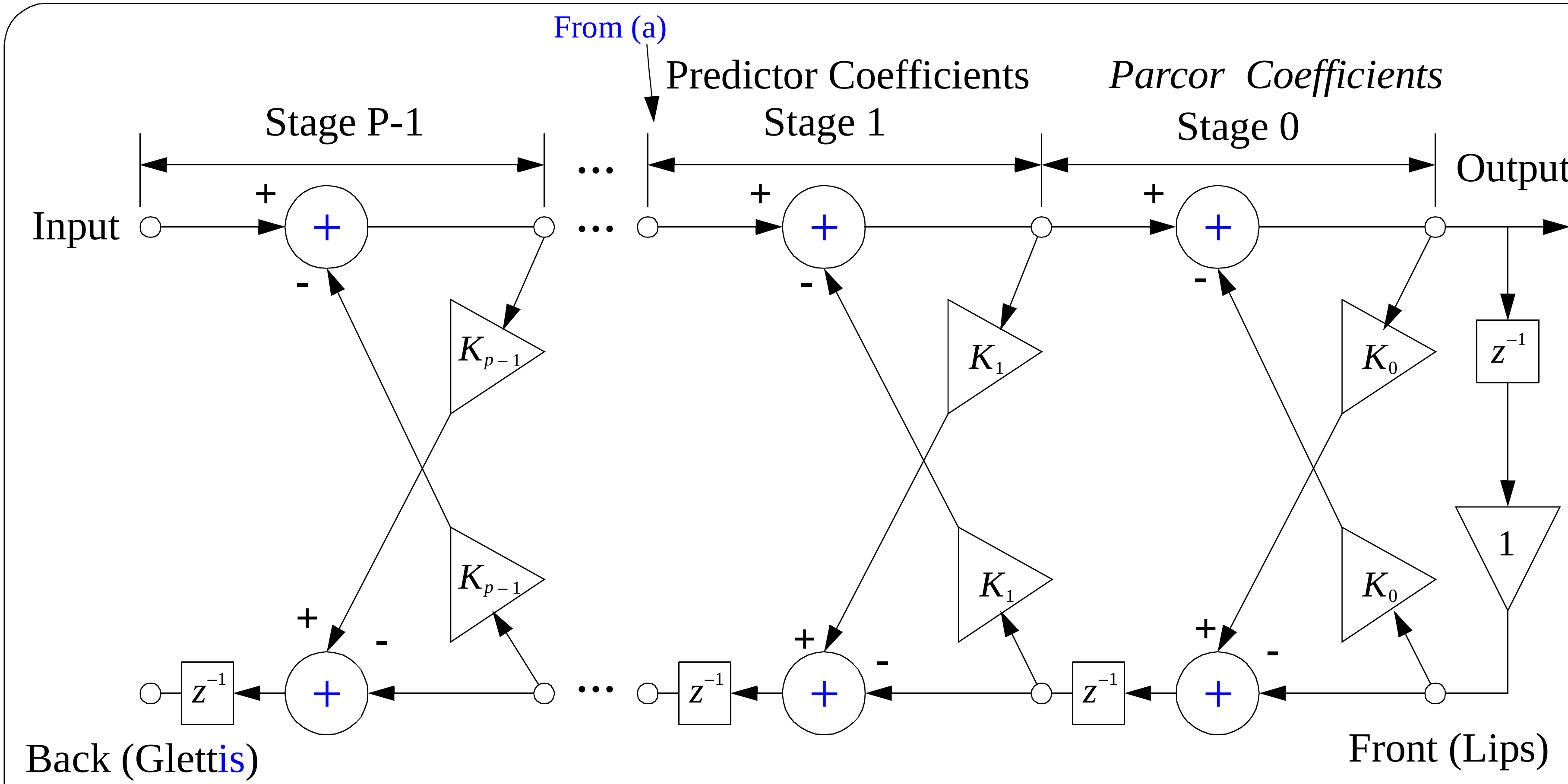


(a) Direct-Form Digital Filter with Variable “a” Coefficients



(b) Acoustic Tube with Variable Area Functions

Figure 29.5 : Two configurations for all pole synthesizers based on LPC analysis.(cont.)



(c) All-Pole lattice Network with Variable "k" Parameters

Figure 29.5 : Two configurations for all pole synthesizers based on LPC analysis.

Synthesis Strategy for Voiced Sounds For unvoiced Sounds.

Vector quantization of the PARCOR Coefficients.

Parameter	Coding	
	Typical 2400-Bit-Per-Second Linear Predictive Encoder	600-Bit-Per-Second Voice Digitizer
Frame rate	44.444Hz	40Hz
Vocal-tract-filter parameters	40 bits/frame	7 bits/frame
Excitation parameters		
Voice/unvoice decision	1 bit/frame	1 bit/frame
Amplitude	6 bits/frame	4 bits/frame
Pitch	6 bits/frame	5 bits/double frame
Synthronization	1 bit/frame	1 bit/double frame
Total number of bits	54 bits/frame	30 bits/double frame

Figure 32.4 : Parameter Coding for 600bps Voice Digitizer

<u>V.Q.Algonithm</u>

* Storage of the patterns.

Compare incoming pattern with All previously stored patterns.

15

2^{N} patterns stored.

* Receiver <u>also</u> has stored availavle.

Send the <u>address</u> of the stored pattern nearest to the actual pattern.

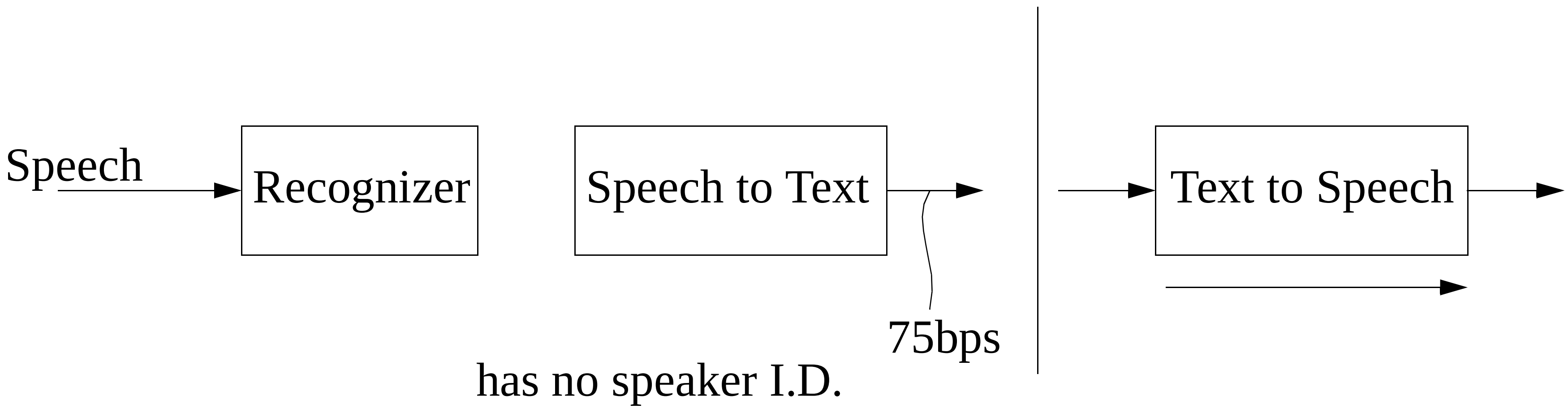
Kang-Coulter 600bps Vocoder

Started with 2400bps —— LPC 10 Vocal Tract Parameter

Reduce # of parameter to 4 or 5

Vector quantization

Frame-fill.



Low Rate Vocoder

Date Rates	*Applications & Performance*	*Algorithms*
30,000 bps	High Fidelity (Telephony)	ADPCM DPCM
10,000-20,000 bps	Medium quality Good Noise Some Noise Immunity	Split Band Systems (Part Modelling, Part Waveform Coding)
5,000-10,000 bps	Reasonable Quality [High Intelligibits]	Voice Excited Vocoders Channel LPC hommerphic
2,000-5,000 bps 2,400 bps (standard)	More Vulnerable to Environment. Speecher I.D.? Secrecy. Pitch Errors.	CELP, Multipolars, STC Channel Vocoders, LPC, Cepstral. Speech Digitization Algorithms.
1,000-2,000 bps	Secrecy. Very Restricted Channels.	Frame-Fill Transformation other Systems.
500-1000 bps	Extremely Restrected or Busy Channels.	Vector Quantization
100-500 bps	Underwater	Recognition-Synthesis Phanetic Vocoder.

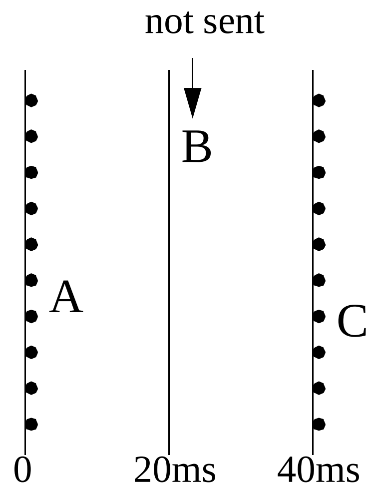
1000-2000bps Frame-fill

start with a complete System —— 2400bps

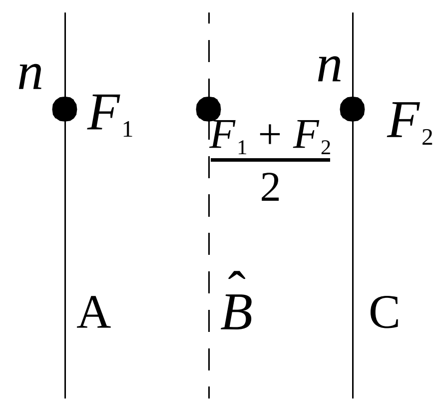
400bps → Excitation

→ 2000bps → Vocal Tract

Analyzer



Synthesizer



Is $B \approx A$?

Or is $B \approx C$

2 Control Bits

Or is neither true.

2400bps → 1200bps

If B is close to A, $B = A$.

If B is close to C, $B = C$.

If neither is true, - Interpolate.

Reduction of # of Parameters

500-10Mbps

C.P. Smith → Pattern Matching

2400bps Vocoder

2000bps Spectrum

50Hz Rate → 40 Bits Per Frame $\underline{\underline{50 \times 40 = 2000}}$

2^{40} Different Spectrum [Channel Vocoder]

2^{20} 1 million

↓

Vector Quantization → LPC

Frame-fill for LPC ?

Predictor coefficients a's $\leftrightarrow$

K-Parameter [PARCOR]

Area Retios.

Log Area Ratios.

] Sensitivity