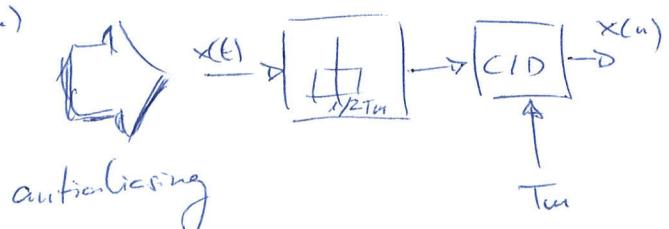
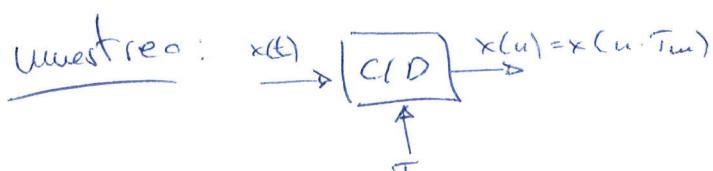
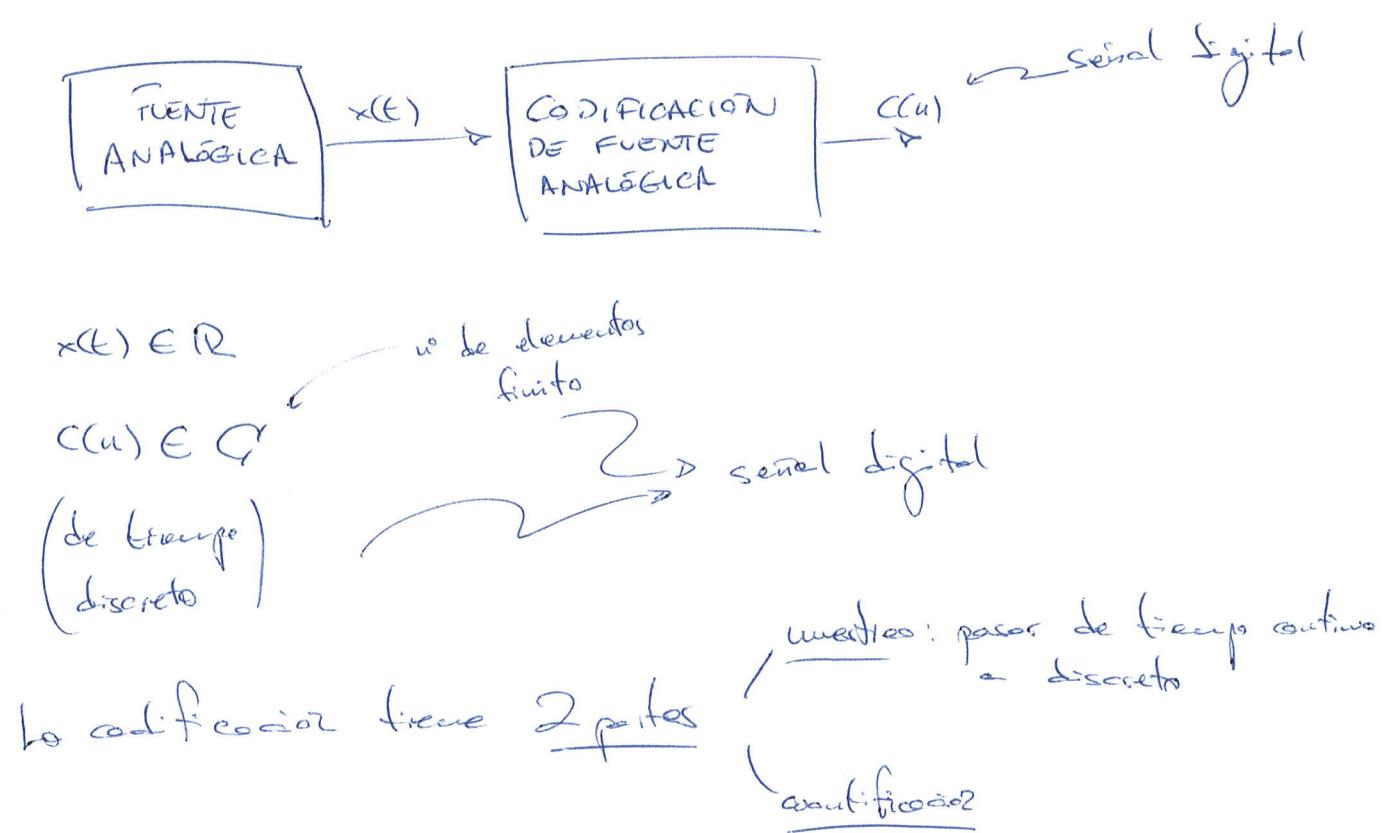


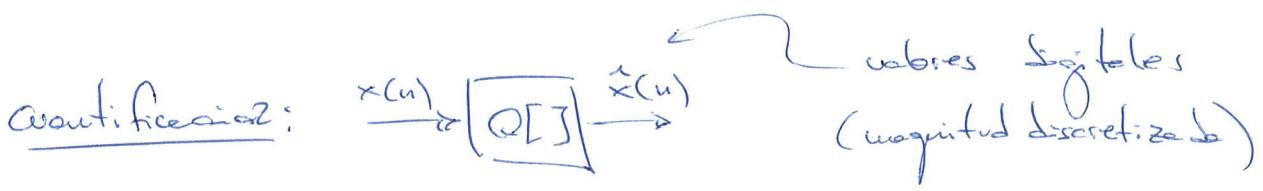
# TAREA 3º: CODIFICACIÓN DE VOZ

## I.- INTRODUCCIÓN A LA CODIFICACIÓN DE FUENTE ANALÓGICA

- Fuente analógica: origen de una señal de información, de naturaleza analógica (de tiempo y magnitud continuas)



$x(t)$  de banda limitada,  $f_m \geq 2B \Rightarrow$  proceso invertible



este proceso implica pérdida de información, por lo que no es totalmente invertible.

$x(u)$  → cantidad de información por muestra infinita  
 $\hat{x}(u)$  → cantidad de información por muestra finita  
 $\Rightarrow$  pérdida de información infinita

No importa, y es hasta sensate hacer la cuantificación, y se para el destinatario puede hacerse transparente.

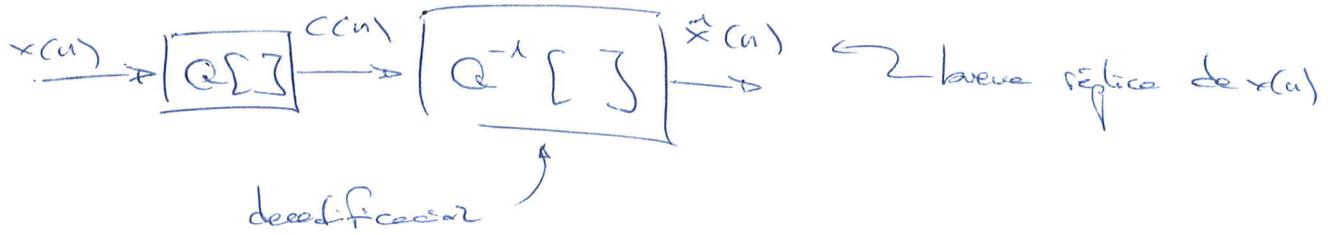
Además, ningún canal de comunicación puede transmitir una cantidad de información infinita

#### - TIPOS DE CODIFICACIÓN DE FUENTE ANALÓGICA:

- de forma de onda
  - escalar
  - vectorial
- parámetricas
  - escalar
  - vectorial

## J.- Codificadores de formas de auto:

Intentar preservar la forma de audio de la señal. En el proceso de decodificación se intenta recuperar ese forma de señal.



Criterio de calidad: error de cuantificación o codificación (ALIA "ruido de cuantificación")

$$|g(n) - x(n)|$$

$\Rightarrow$  potencia de la señal de error de cuantificación

Para medir la calidad podemos usar el cociente:

$$\frac{S}{N} = \frac{\text{Potencia}(x(n))}{\text{Potencia}(g(n))}$$

Esto se puede hacer de 2 formas / escalar  
(vectorial)

- escalar: cuantificar cada muestra independientemente
- vectorial: cuantificar bloques de muestras. Necesita menos bits/muestra pero une misma SNR, aunque es más complejo

2.- Codificaciones paramétricas: se intenta que el destinatario

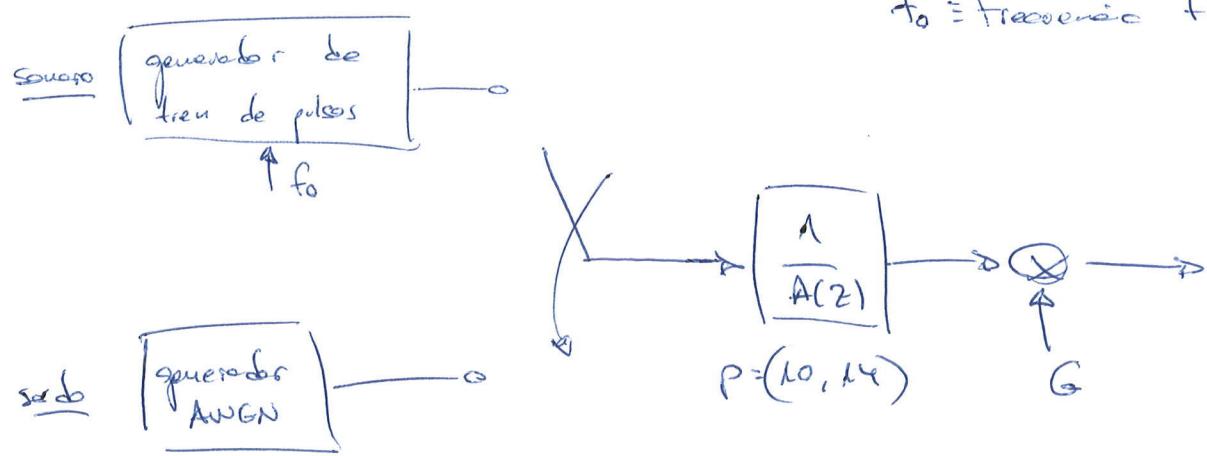
"entienda". Por ejemplo, el oído es insensible a los cambios de fase.

$\Rightarrow$  no hay medida de calidadérica, debe hacerse empíricamente.

Se trate de que el codificador genere el modelo de la señal  
 $\Rightarrow$  obtener los parámetros del modelo.

En el destino se usan los parámetros para sintetizar la señal

## - LPC (Codificador de Predicción Lineal)



### • Codificación (análisis):

1.- Sequestrar la voz ( $\sim 20$  ms)

2.- Determinar tipo de sonido (silencio/silencio/silencio) y codificarlo

3.- Si es sonido, estimar  $f_0 \rightarrow \hat{f}_0$

4.- Si es sonido/silencio, estimar  $A(z)$  (medir AR)  $\Rightarrow$  coeficientes PARCOR  $\rightarrow \hat{k}_i$

y como normalmente se canaliza, estimar la ganancia  $G \rightarrow \hat{G}$

desarrollar a codificar: - tipo de señal: 2 bits  
-  $\hat{f}_0$   
-  $\hat{k}_i$   
-  $\hat{G}$

### • Síntesis:

$\rightarrow$  Ajustar los parámetros en los tramos de datos los datos identificados se tienen que ver con la forma de la señal.

✗ se pierde información

✓ se aumenta la infertilidad

✓ hacen falta menos bits

Usos: militares

- Codificadores mixtos { forma de onda + parámetros  
⇒ ✓ mejoran la calidad

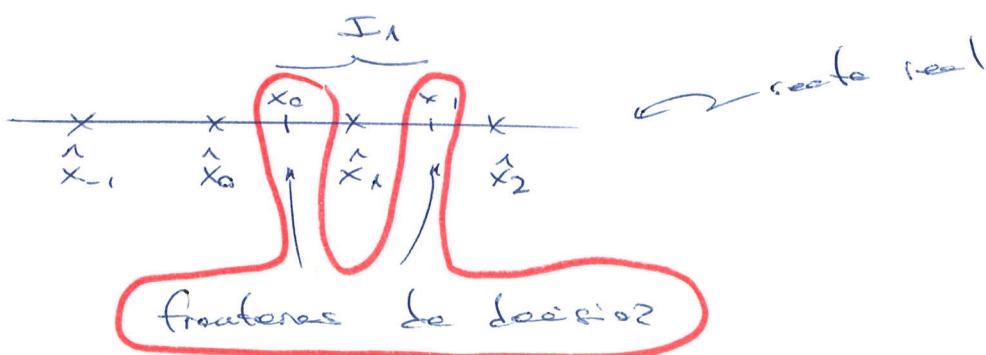
Usos: teléfono móvil

## 2.- CUANTIFICACIÓN ESCALAR

Se puede aplicar a una señal  $x(n)$  o a un conjunto de parámetros. Eliminemos el índice:  $x(n) \rightarrow x$   
 $\hat{x}(n) \rightarrow \hat{x}$



$$q = x - \hat{x} = x - Q\{x\}$$



El conjunto de valores  $\hat{x}_i$  se llaman niveles de representación o reconstrucción.

Minimizar el error de cuantificación:

$$x \rightarrow \hat{x}_k \text{ tal que } |x - \hat{x}_k| \leq |x - \hat{x}_i| + i$$

$x_0, x_1, \dots$  = niveles o umbrales de decisión que definen los intervalos  $I_i$

$$\boxed{x \in I_k \Rightarrow Q[x] = \hat{x}_k}$$

$$\begin{aligned} I_k &= (x_{k-1}, x_k) \\ &= \left( \frac{\hat{x}_{k-1} + \hat{x}_k}{2}, \frac{\hat{x}_k + \hat{x}_k}{2} \right) \end{aligned}$$

$$x_k = \frac{\hat{x}_k + \hat{x}_{k+1}}{2} \quad \text{= valor óptimo para minimizar el error de cuantificación}$$

- Corrección estadística del cuantificador:

Particulares de:

- fdp de  $x$  conocida:  $f_x(x)$
- $E[x] = 0$  si no, se ajusta la media

$$P_k = Q_k^2 = \int_{-\infty}^{\infty} x^2 f_x(x) dx \quad \begin{array}{l} \text{para determinar la anchura} \\ \text{de los niveles de representación} \end{array}$$

• Potencia del ruido de cuantificación:

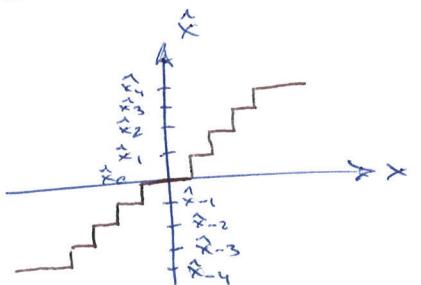
$$P_q = \int_{-\infty}^{\infty} (x - Q[x])^2 f_x(x) dx$$

• Média del ruido de cuantificación:

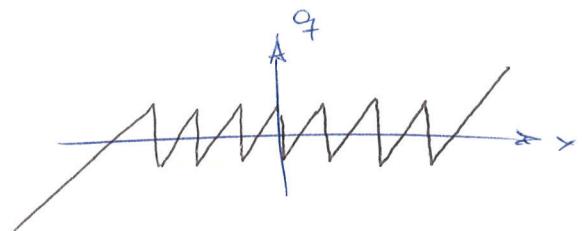
$$m_q = E[q] = \int_{-\infty}^{\infty} (x - Q[x]) f_x(x) dx$$

En la mayoría de los casos,  $\omega_g \approx g \Rightarrow P_g = \sigma_g^2$

- fusión de cuantificación:



- fusión de errores:



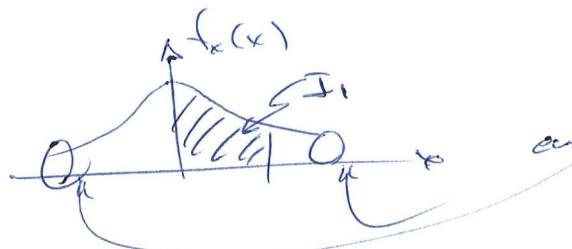
fdp del ruido de cuantificación:

$$q = x - Q[x] = g(x)$$

$$f_Q(q) = \sum_{x=g^{-1}(q)} \frac{f_x(x)}{g'(x)}$$

$$\text{con } g'(x) = 1$$

$$f_Q(q) = \sum_{x=g^{-1}(q)} f_x(x)$$



los extremos no logran el límite en el error  $\Rightarrow$  saturación

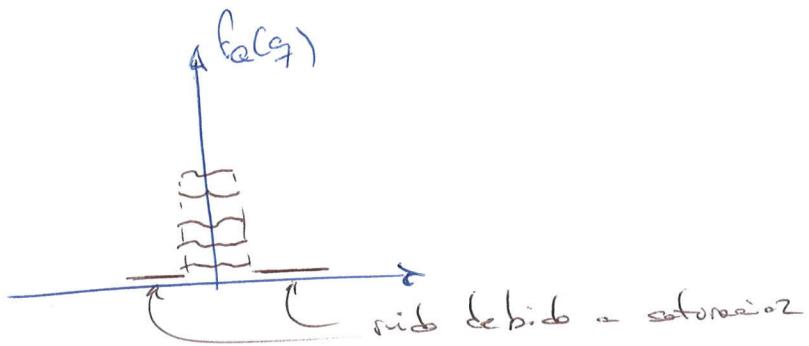
definimos 2 funciones

$$f_x^{(k)}(x) = \begin{cases} f_x(x) & x \in I_k \\ 0 & x \notin I_k \end{cases}$$

$$f_x(x) = \sum_k f_x^{(k)}(x)$$

$$\Rightarrow f_Q(q) = \sum_k f_x^{(k)}(q + \hat{x}_k)$$

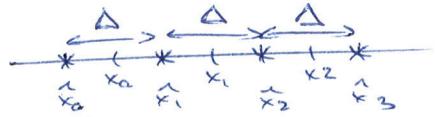
Este se usa en cuantificadores de paso niveles. Para niveles niveles,  $f_Q(q)$  tiene formas cuadráticas (ej: uniforme)



## 2.1.- COUPLING UNIFORME

Muy usada en electrónica. Coupling con intervalos del mismo tamaño  
 $\Rightarrow$  niveles de condensador equiespaciados.

parámetro determinante: número de niveles  
 (suele ser potencia de 2)



2 tipos  
 (simétricos)

(\*) 3.1(\*)

/ mid-riser  $\rightarrow$  No hay valor para el cero ( $x_0=0$ )  
 completamente simétricos

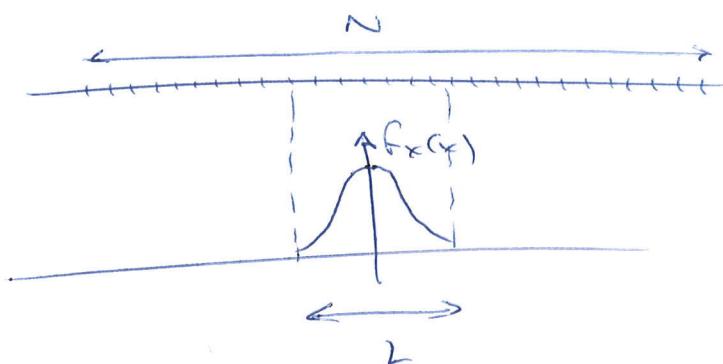
\ mid-trad  $\rightarrow$  Si hay nivel para el cero ( $x_0=0$ )  
 hay un nivel vacío en la parte  
 positiva (o negativa, niveles fijos)  
 para un número par de niveles

Cuando se codifica silencio (nivel muy bajo), resultados  
 muy distintos según tipo:

- mid-riser: oscil entre los 2 niveles inferiores
- mid-trad: casi siempre nivel 0.

- Consideraciones acerca del número de niveles:

cuantificados:



señal:

Si el nivel de señal de entrada es pequeño, se desaprovechan los niveles del cuantificador.

Si hay pocos niveles utilizados ( $L$ ), el error de cuantificación tiene relación estadística con la señal original:

$g(u)$  correlada con  $x(u)$   
 $g(u)$  con d.e.p. poco bajo  $\Rightarrow$  correlada  
no es fácil obtener  $f_{\bar{g}}(q)$   $\rightarrow$  usar fórmulas  
genéricas (integrandos, ...)

3: hay **muchos** niveles utilizados:

$g(u)$  correlada con  $x(u)$   
 $g(u)$  auto-correlación casi nula:  $\mathbb{E}[g(u)g(u)] = \mathbb{E}[g(u)]\mathbb{E}[g(u)]$   
 $f_g(q)$  constante (tiende a ser uniforme en el intervalo)

El error varía entre  $-\frac{\Delta}{2}$  y  $\frac{\Delta}{2}$ .

$$\Rightarrow \left[ f_g(q) = \frac{1}{\Delta} \quad q \in \left[ -\frac{\Delta}{2}, \frac{\Delta}{2} \right] \right] \text{ Lf}$$

Ejemplo: (\* 3.2 \*, Fig 5.11)

$$\text{Caus. de ruido } L^A : \frac{S}{N} = \frac{\sigma_x^2}{\sigma_q^2}$$

$$\boxed{G_q^2 = \int_{-\Delta/2}^{\Delta/2} q^2 f_q(q) dq = \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} q^2 dq = \frac{\Delta^2}{12}}$$

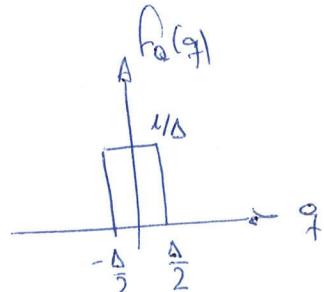
$$\left. \begin{array}{l} \Delta = \frac{2x_m}{N} \\ N = 2^B \end{array} \right\} \begin{array}{l} \Delta \equiv \text{longitud del intervalo de quantificación} \\ N \equiv \text{niveles} \\ B \equiv \text{número de bits de quantificación} \end{array}$$

$$\left. \begin{array}{l} N \equiv \text{número de niveles de representación} \\ L \equiv \text{número de niveles de representación útiles} \end{array} \right\}$$

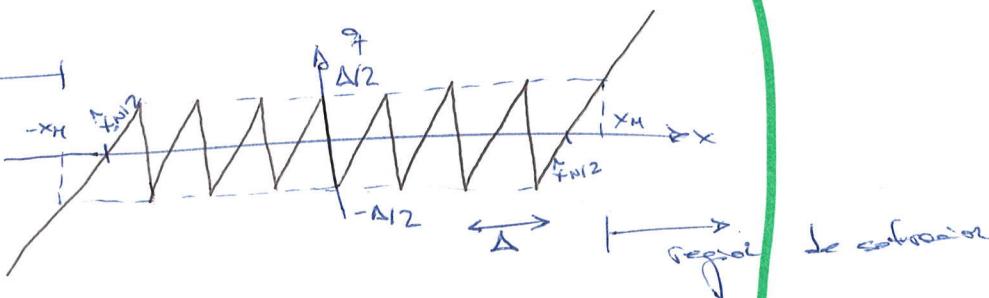
Si  $L > 10$  y no hay saturación

Potencia del ruido de quantificación:

$$P_q^2 = \sigma_q^2 = \frac{\Delta^2}{12}$$



región de saturación



SNR de quantificación:

$$\boxed{S/N = \frac{G_x^2}{P_q^2} = 12 \frac{\sigma_x^2}{\Delta^2} = \frac{12}{4} \frac{\sigma_x^2 N^2}{x_m^2} = 3 \cdot 2^B \frac{G_x^2}{x_m^2}}$$

$$\Delta = \frac{2x_m}{N}$$

$$N = 2^B$$

$$S/N[\text{dB}] = 10 \log S/N = 4,8 + 6B - 20 \log \frac{x_m}{\sigma_x} \quad (\text{sin saturación})$$

$$\frac{x_m}{\sigma_x} = f_L \equiv \underline{\text{factor de sobreexpos}}.$$

Ruido granular:

$$N_g = \int_{-x_m}^{x_0} (x - \bar{x}_0)^2 f_x(x) dx + \sum_{k=1}^{N-2} \int_{x_{k-1}}^{x_k} (x - \bar{x}_k)^2 f_x(x) dx + \int_{x_{N-2}}^{x_m} (x - \bar{x}_{N-1})^2 f_x(x) dx$$

Ruido total:

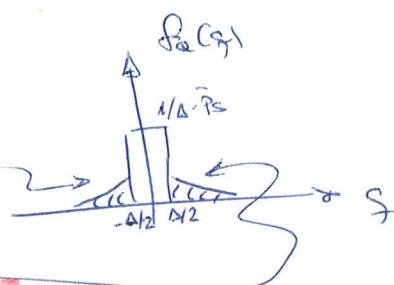
$$G_g^2 = N_g + \int_{-\infty}^{-x_m} (x - \bar{x}_0)^2 f_x(x) dx + \int_{x_m}^{\infty} (x - \bar{x}_{N-1})^2 f_x(x) dx$$

$N_g \equiv$  ruido de fondo

Para el ruido granular podemos evaluar las integrales si se tiene un número de niveles suficiente (3 o más bits)

$$G_g^2 = \bar{P}_S \frac{\Delta^2}{\lambda^2} + N_g$$

$\bar{P}_S \equiv P_S \equiv$  probabilidad de saturación



$$\text{Además: } \bar{x}_0 = -x_m + \frac{\Delta}{2}$$

$$\bar{x}_{N-1} = x_m - \frac{\Delta}{2}$$

$$G_g^2 = \int_{-\infty}^{x_0} (x - \bar{x}_0)^2 f_x(x) dx + \sum_{k=1}^{N-2} \int_{x_{k-1}}^{x_k} (x - \bar{x}_k)^2 f_x(x) dx + \int_{x_{N-2}}^{\infty} (x - \bar{x}_{N-1})^2 f_x(x) dx$$

## 2.2.- CANTIFICACIÓN ÓPTIMA

- datos  $\left\{ \begin{array}{l} f_x(x) \\ N \end{array} \right.$
- parámetros a calcular  $\left\{ \begin{array}{l} x_k \\ x_k \\ x_k \end{array} \right. \quad \left\{ \begin{array}{l} k=0: N-1 \\ k=0: N-2 \end{array} \right. \quad \left\{ \begin{array}{l} 2N-1 \text{ parámetros} \end{array} \right.$

Es un problema de minimización sin solución analítica  
(traspaso que no tienen)

$$1) \frac{\partial}{\partial x_i} \hat{\sigma}_q^2 = f_x(x_i)(x_i - \hat{x}_i)^2 - f_x(x_i)(x_i - \hat{x}_{i+1}) = 0$$

$$\Rightarrow x_i = \frac{1}{2}(x_i + x_{i+1}) \quad [1]$$

Quedan sobre  $N$  variables,  $\hat{x}_i \quad \left\{ \begin{array}{l} x_0 = -\infty \\ x_N = \infty \end{array} \right.$

$$2) \frac{\partial}{\partial x_i} \hat{\sigma}_q^2 = - \int_{x_{i-1}}^{x_i} 2(x - \hat{x}_i) f_x(x) dx = 0$$

$$x_i = \frac{\int_{x_{i-1}}^{x_i} x f_x(x) dx}{\int_{x_{i-1}}^{x_i} f_x(x) dx} \quad [2] \quad \left( \begin{array}{l} \text{centroide de la región} \\ x_{i-1} \leq x \leq x_i \end{array} \right)$$

[1] y [2] son las condiciones de Lloyd-Max

$\hat{x}_i$ : se llaman soluciones analíticas

Problemas del cantificador óptimo:

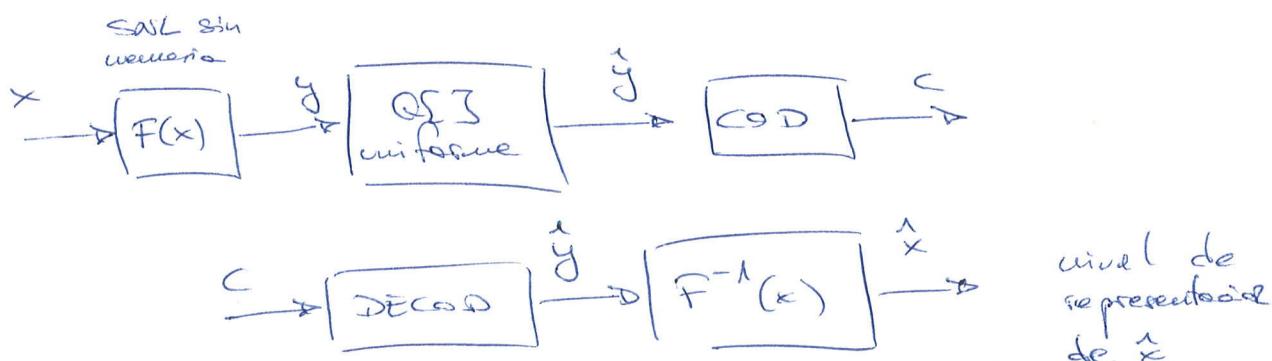
x es muy sensible al nivel de potencia. Se puede solucionar con un CAG.

x es la señal no es estacionaria, un cantificador óptimo es una inversor ya que tiene salto en cada instante de tiempo

Ejemplo resuelto: (\* 3.5\*)

[Para una fdp uniforme, el acutificador óptimo será el uniforme]

## 2.3.- CUANTIFICACIÓN NO UNIFORME



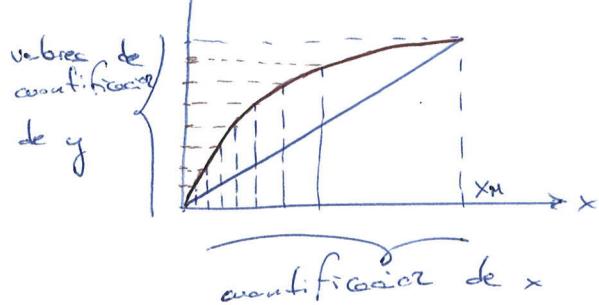
Uno de los más prácticos es el acutificador logarítmico

Buscar que las preferencias sean inseparables = la preferencia de la señal de entrada.

$$\left. \begin{array}{l} \text{uniforme} \rightarrow S/N = k \cdot \sigma_x^2 \\ \text{óptimo} \rightarrow S/N \text{ muy sensible a } \sigma_x^2 \end{array} \right\}$$

Se prede hacer que  $S/N$  sea poco dependiente de  $\sigma_x^2$  dentro de un rango de entrada

las funciones no lineales deben ser caóticas, es decir, deben preservar el signo y el rango



$$\hat{y}_k = F(\hat{x}_k)$$

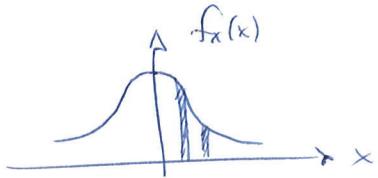
Fuera de compresión

## Ruido de cuantificación de un cuantificador no uniforme:

Suposiciones:

a)  $N$  es tan grande que  $f_x(x) \approx \text{cte}$   $\forall x \in I_k$

(para todos los intervalos de decisión)



b) No hay saturación:  $|x| \leq x_{\text{re}}$

$$f_x(x) = f_x(\hat{x}_k) \quad x \in I_k$$

$$\Delta_k = x_k - x_{k-1} \quad \text{tamaño del intervalo } I_k$$

Probabilidad de ocurrencia del nivel  $\hat{x}_k$ :

$$P_k = \Delta_k f_x(\hat{x}_k) \quad \sum_k P_k = 1$$

$$\sigma_q^2 = \sum_k \int_{x_{k-1}}^{x_k} (x - \hat{x}_k)^2 f_x(x) dx = \sum_k \int_{x_{k-1}}^{x_k} (x - \hat{x}_k)^2 \frac{P_k}{\Delta_k} dx =$$

$$= \sum_k \int_{-\frac{\Delta_k}{2}}^{\frac{\Delta_k}{2}} x^2 \frac{P_k}{\Delta_k} dx = \sum_k \frac{1}{12} P_k \Delta_k^2$$

Es una generalización del ruido de cuantificación para el caso no uniforme

Lo relacionamos con  $F(x)$ :

$$F'(\hat{x}_k) = \left. \frac{dy}{dx} \right|_{x=\hat{x}_k} \approx \frac{y_k - y_{k-1}}{x_k - x_{k-1}} \quad \text{uniforme} = \Delta$$

$$\sum_k \left[ F'(\hat{x}_k) = \frac{\Delta}{\Delta_k} \right]$$

$$\sigma_g^2 = \frac{1}{12} \sum_k p_k \frac{\Delta^2}{(F'(x_k))^2} = \frac{\Delta^2}{12} \sum_k p_k \frac{1}{(F'(x_k))^2}$$

Otra aproximación:

- 1. intervalos muy pequeños
- 2. sin saturación

$$\left[ \sigma_g^2 = \frac{\Delta^2}{12} \int_{-x_H}^{x_H} \frac{f_x(x)}{(F'(x))^2} dx \right] \quad S/N = \frac{\sigma_x^2}{\sigma_g^2} = \frac{12 \int_{-x_H}^{x_H} x^2 f_x(x) dx}{\Delta^2 \int_{-x_H}^{x_H} \frac{f_x(x)}{(F'(x))^2} dx}$$

Objetivo:  $S/N = \text{cte}$  e independiente del nivel de señal

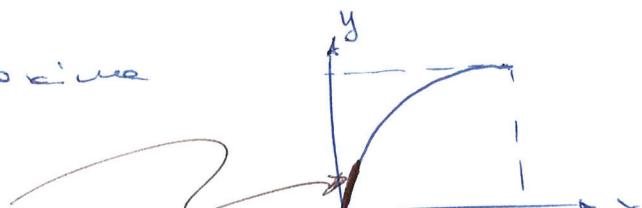
$$\Rightarrow \boxed{F'(x) = \frac{K}{x}} \quad \Rightarrow \boxed{F(x) = k \ln x + k_0}$$

CUANTIFICADOR LOGARÍTMICO

$x=0 \Rightarrow F(x) = -\infty$  !  $\rightarrow$  no se puede aplicar a todos los valores de  $x$

Para  $x \downarrow 0$  se approxima

por una recta



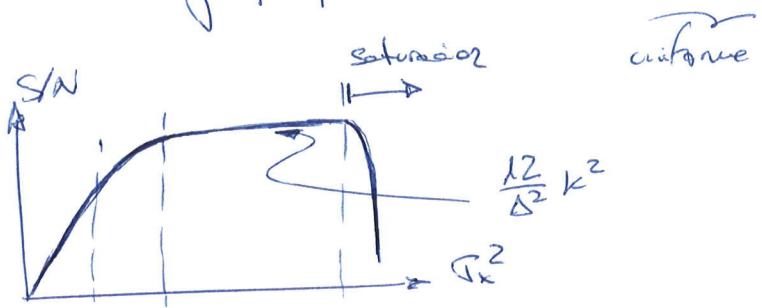
- CARACTERÍSTICAS DEL CUANTIFICADOR LOGARÍTMICO:

1.- En la zona en que es constante:  $|S/N = \frac{12}{\Delta^2} K^2|$

2- Generación de cuantizado:

para  $x \downarrow 0 \Rightarrow \left( G_C = \lim_{x \downarrow 0} \frac{dF(x)}{dx} = F'(0) \right)$

Señal muy pequeña :  $S/N = (S/N)_u \cdot G_c^2$



3.-  $F(x)$  tiene simetría impar

- FUNCIONES DE COMPRENSIÓN PRÁCTICAS:

1.- Ley  $\mu$ : estándar de la telefonía digital analógica

$$\mu = 255$$

$$f(x) = x_H \frac{\ln(1 + \mu \frac{|x|}{x_H})}{\ln(1 + \mu)} \operatorname{sign}(x)$$

$$\begin{pmatrix} f(0) = 0 \\ f(x_H) = x_H \end{pmatrix} \times \text{bb} \Rightarrow f(x) \approx \frac{x_H}{\ln(1 + \mu)} \mu \frac{|x|}{x_H} = \frac{\mu}{\ln(1 + \mu)} |x| \operatorname{sign}(x)$$

$$= \frac{\mu}{\ln(1 + \mu)} x$$

$$G_c = \frac{\mu}{\ln(1 + \mu)}$$

$$\times \dagger \dagger \Rightarrow \ln\left(1 + \mu \frac{|x|}{x_H}\right) \approx \ln \mu \frac{|x|}{x_H}$$

$$R = \frac{x_H}{\ln(1 + \mu)}$$

$$S/N = \frac{12}{\Delta^2} k^2 = 3 \cdot 2^{2B} \frac{1}{(\mu^2 (1 + \mu))}$$

$\Delta = \frac{2x_H}{2^B}$

región constante

$$\mu = 255 \Rightarrow \begin{cases} G_C = 33,3 \text{ dB} \\ S/N = 6B - 10 \text{ dB} \end{cases}$$

2.- Ley A: estandar europeo, normalizado por la ITU

$$A = 87,56$$

$$F(x) = \begin{cases} \frac{A \cdot |x|}{1 + \ln A} \operatorname{sign}(x) & 0 \leq |x| \leq \frac{x_M}{A} \\ x_M \frac{1 + \ln \left( \frac{A \cdot |x|}{x_M} \right)}{1 + \ln A} \operatorname{sign}(x) & \frac{x_M}{A} \leq |x| \leq x_M \end{cases}$$

$$G_C = \frac{A}{1 + \ln A}$$

$$k = \frac{x_M}{1 + \ln A}$$

$$S/N = \frac{3 \cdot 2^{2B}}{(1 + \ln A)^2}$$

región constante

$$A = 87,56 \Rightarrow \begin{cases} G_C = 24 \text{ dB} \\ S/N = 6B - 10 \text{ dB} \end{cases}$$

(\* 3.3 \*)  $\rightarrow$  Fig 5.15: evaluar la curva de compresión ojo

(\* 3.4 \*)  $\rightarrow$  prestaciones

$$B = 7 \rightarrow \begin{cases} \mu = 100 \Rightarrow SNR = 33 \text{ dB} \\ \mu = 500 \Rightarrow SNR = 31 \text{ dB} \end{cases}$$

$\mu \uparrow \Rightarrow$  parte plana más larga

## 2.4.- ESTÁNDAR PCM (PULSE CODE MODULATION)

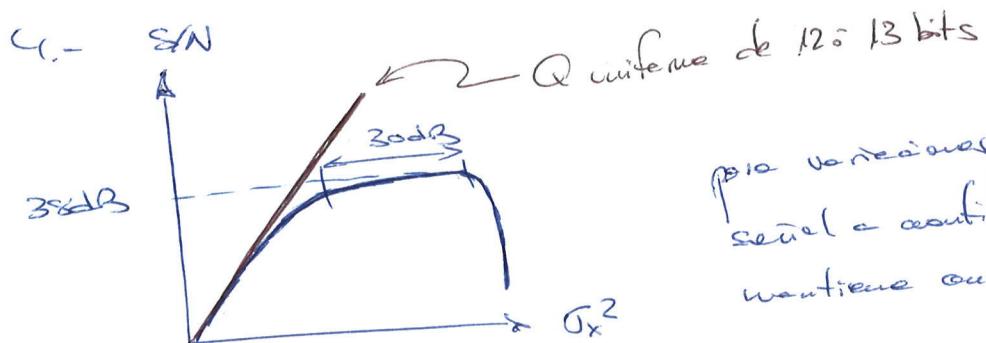
Quantificador logarítmico usado en teléfono digital

### - CARACTERÍSTICAS:

1.- Ancho de banda para la voz en teléfono: 300-3300 Hz

2.-  $f_m = 8 \text{ kHz}$       Quantificador logarítmico de 8 bits       $\Rightarrow v = 64 \text{ kbps}$

3.- EEUU.  $\rightarrow$  ley  $\mu$  ( $\mu = 255$ )  
ITU  $\rightarrow$  ley A ( $A = 87,56$ )

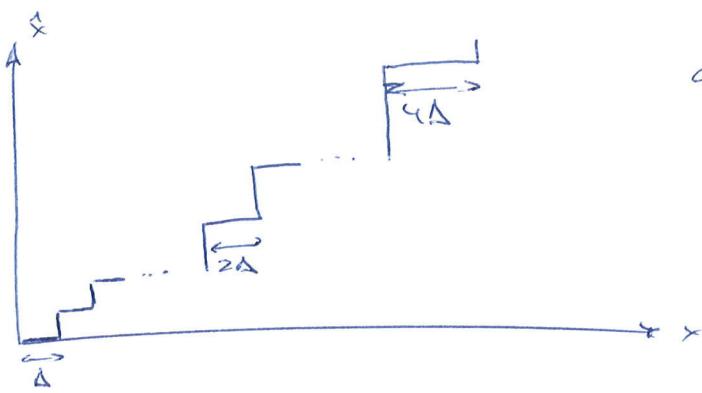


Este es similar a las leyes A y  $\mu$

Para valores pequeños de la señal se comparten con un cuantificador uniforme de 12 o 13 bits

5.- No se implementa la función de la ley  $\mu$  o A, sino una aproximación por tramos sucesivos (tramos)

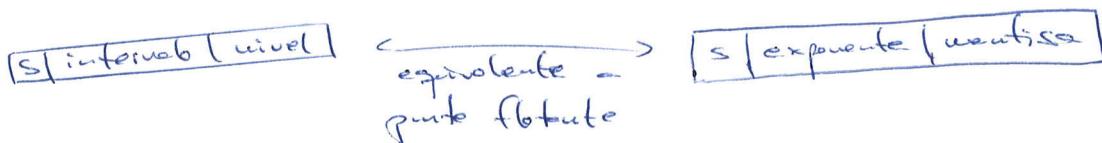
$\mu = 255 \Rightarrow 8$  tramos para valores positivos y 8 para negativos con 16 niveles por cada tramo



al siguiente tramo, el factor de los errores se dobla

## Polarización de 8 bits:

- signo: 1 bit
- intervalo de aproximación (nivel): 3 bits
- nivel dentro del intervalo: 4 bits



## 2.5.- CALIDAD DE LA SEÑAL DE VOZ

- ALTA CALIDAD:  $B = 7-10 \text{ kHz}, 64 \text{ kbps}$

- TELEFÓNICA:  $B = 3,4 \text{ kHz}, 48-64 \text{ kbps}$

- COMUNICACIONES MILITARES:  $1-4,8 \text{ kbps}$  (solo se requiere inteligibilidad)

- SINTÉTICA: se extrae la información del mensaje vocal, extrayendo el texto. Se puede producir una calidad muy alta, aunque este totalmente decorrelado con el bájico.

Vocación:  $\sim 200 \text{ bps}$  (velocidad a la que habla el bájico)

CARACTERIZACIÓN DE LA SEÑAL DE VOZ: (\*3.7 - 3.8\*)

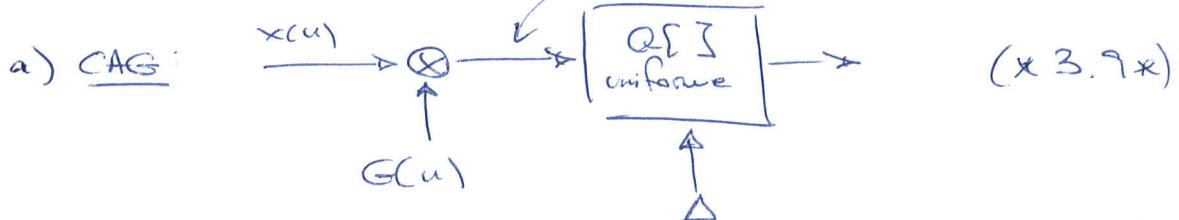
En la práctica se suele usar  $f_L = \frac{x_m}{\Delta x} = 4$

que da una probabilidad de solamente de  $p_s < 0,04$

## 2.6 - CUANTIFICACIÓN ADAPTATIVA

Se sol no estacionario  $\Rightarrow$  cuantificadores sensibles a la potencia

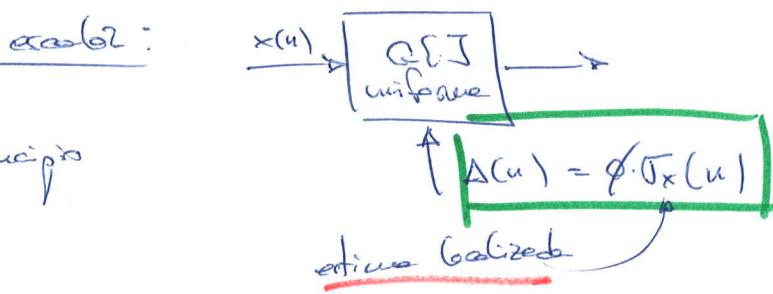
2) insensibilizar la cuantificación de las variaciones temporales de la potencia:



b) Tono variable del escalar:

Consideremos en principio este caso

$$G_x(u) = \sqrt{\sigma_x^2(u)}$$



$$\sigma_x^2(u) = \sum_{m=1}^L x^2(m) w(u-m)$$

energía localizada

Ventana rectangular de longitud L:  $w(u) = \frac{1}{L}; u=1:L$

$$G_x^2(u) = \frac{1}{L} \sum_{m=1}^L x^2(u-m)$$

$\Rightarrow$  potencia

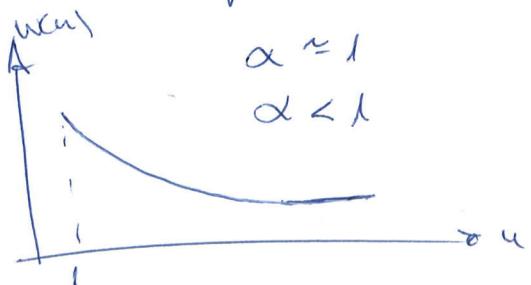
$\sigma_x^2(u)$  varía con el tiempo

$\Rightarrow A(u)$  también

$\Rightarrow L \downarrow \Rightarrow \sigma_x(u)$  varía más rápido

Otro tipo de ventana para mayor eficiencia adaptativa es  
la ventana exponencial:

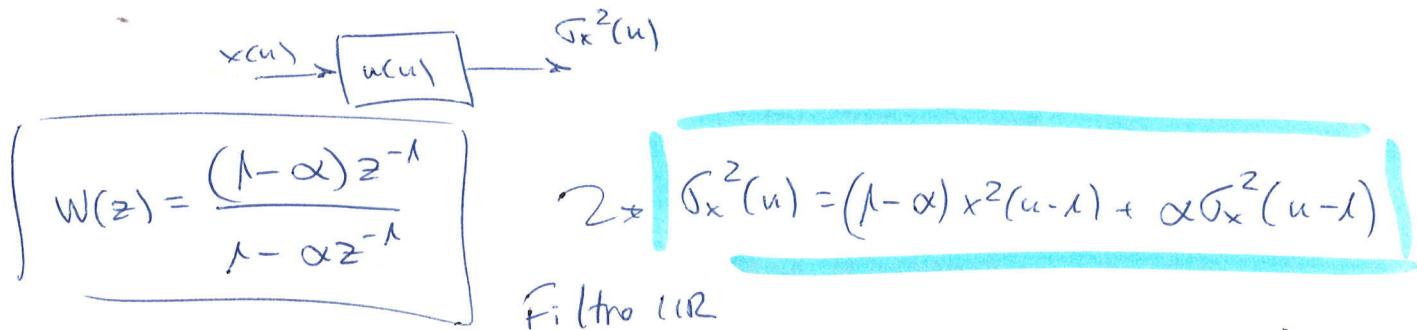
$$w(u) = \begin{cases} (1-\alpha)\alpha^{u-1} & u \geq 1 \\ 0 & u \leq 0 \end{cases}$$



$$\sigma_x^2 = (1-\alpha) \sum_{u=1}^{\infty} x^2(u-u) \alpha^u$$

Tienen más peso los muestras más recientes.

Nuevamente, es un filtro con  $h(u) = w(u)$



Continua  $\rightarrow z = 1 \Rightarrow$  ganancia unitaria

En los 2 tipos se fijan estos máximos y mínimos:

$$\begin{cases} \Delta_{\max} \\ \Delta_{\min} \end{cases} \quad \begin{cases} G_{\max} \\ G_{\min} \end{cases}$$

A veces, en vez de usar  $e^2(u)$ , se implica tener una raíz:

$$h(u) = \sum_u |x(u)| w(u-u)$$

$$\rightarrow \Delta(u) = \phi \cdot h(u)$$

$G(u) \circ \Delta(u)$  habrá que aplicar la función al desfilar, para obtener la operación inversa

$\Rightarrow$  codificar también  $\Delta(u) \circ G(u)$

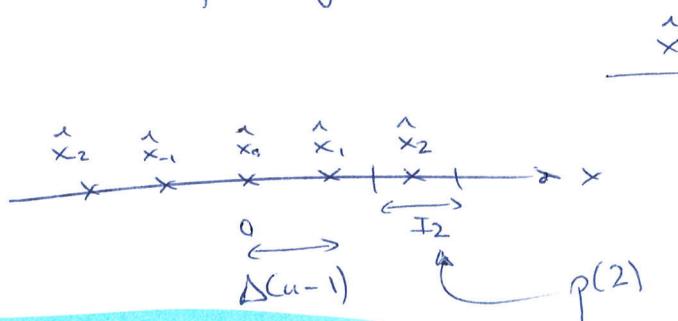
Hay otros 2 tipos:

- Adaptar hacia adelante ("feed-forward", o directo): estimar  $\hat{x}_k$  a partir de los señales de entrada ( $* 3.6 *$ )
- Adaptar hacia atrás ("feedback", retroalimentado): estimar  $\hat{x}_k$  a partir de las muestras codificadas a la salida. Así, no se precisa transmitir  $G(u)$  ni  $\Delta(u)$ . ( $* 3.19 *$ )  
Se pueden propagar errores.

## 2.7.- CUANTIFICACIÓN ADAPTATIVA REALIMENTADA (SAYANT)

( $* 3.11 *$ )

El tamaño del escalar se modifica instantáneamente al adaptar la cuantificación.

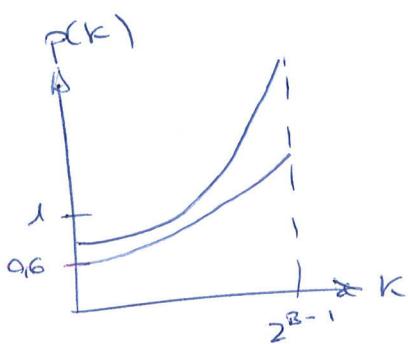


$f$  viene dado por  $\hat{x}_k(u-1)$   
(intervalo en que se encuentra)

$$\Delta(u) = \Delta(u-1) / p(k)$$

Si se cuantifica con niveles altos,

hay que tender a  $\Delta$  alto:



$$k = |c(u-1)| \in [0, 2^{B-1}]$$

- $k \uparrow \Rightarrow p \uparrow$
- $k \downarrow \Rightarrow p \downarrow$

Plátanos extremos  $\Rightarrow \Delta(u) \uparrow$   
Intervalo central  $\Rightarrow \Delta(u) \downarrow$

Esto funciona en promedio, y no necesita valores anteriores de  $x$  ni realimentación

Fixado  $B$ , los posibles multiplicadores son  $p(k)$  que caigan dentro de los límites (Rabiner, pp 20 - 207)

Deben ser neutrales en esos límites (Amin, Durak)  
Para mayor estabilidad

Estos límites también controlan la propagación de errores  
en cierta medida

"Factor de diseño": usar  $\beta \approx 1, \beta < 1$  caso:

$$\Delta(u) = (\Delta(u-1))^{\beta} P\left(\frac{|x|}{\Delta(u-1)}\right)$$

## 2.8.- CUANTIFICACIÓN DIFERENCIAL

B.b. Rabiner 225-232

(\* 3.12 - 3.16 \*)

- Señales con algo de correlación  $\rightarrow$  mejorar las prestaciones del cuantificador aprovechando esa correlación

- Los muestras se parecen mucho:  $x(u) - x(u-1) \downarrow$
- Interesar cuantificar la diferencia entre el valor de la señal y el valor predicho:

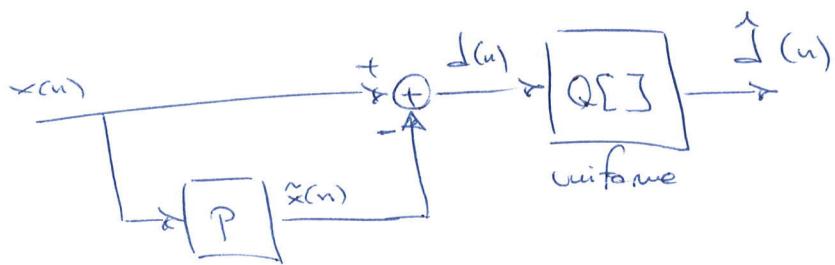
$$d(u) = x(u) - \hat{x}(u)$$

(error de predicción)  
 $\hat{x}(u)$  = predicción

$$\hat{x}(u) = f(x(u-1), \dots, x(u-p))$$

representación lineal:

$$\hat{x}(u) = - \sum_{k=1}^p a_p(k) x(u-k)$$



$$A(z) = 1 + \sum_{k=1}^P c_p(k) z^{-k}$$

Beneficio: reducción del nivel de  $x(n)$

$$e(n) = d(n) - \hat{d}(n) = \underline{\text{error de quantificación}}$$

$$(S/N)_D = \frac{G_x^2}{G_e^2} = \frac{G_x^2}{G_d^2} \cdot \frac{G_d^2}{G_e^2}$$

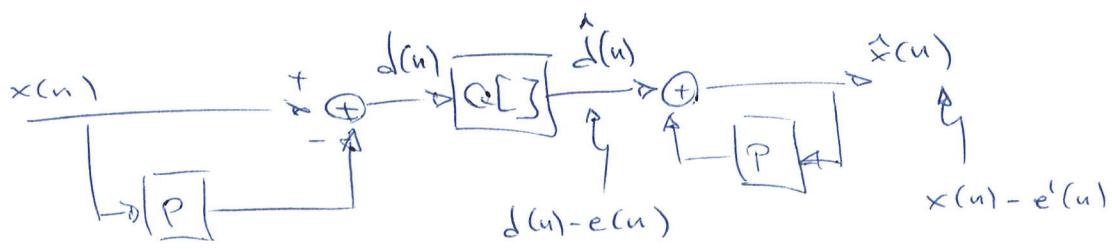
SNR del QΣ uniforme,  $(S/N)_Q$

$$\boxed{(S/N)_D = G_p (S/N)_Q} \quad G_p \geq 1$$

Si este esquema no funciona, y que hay que recuperar la señal



Por tanto, para ver  $\hat{x}(n)$ :



$$\left. \begin{array}{l} e(n) = d(n) - \hat{d}(n) \\ e'(n) = x(n) - \hat{x}(n) \end{array} \right\}$$

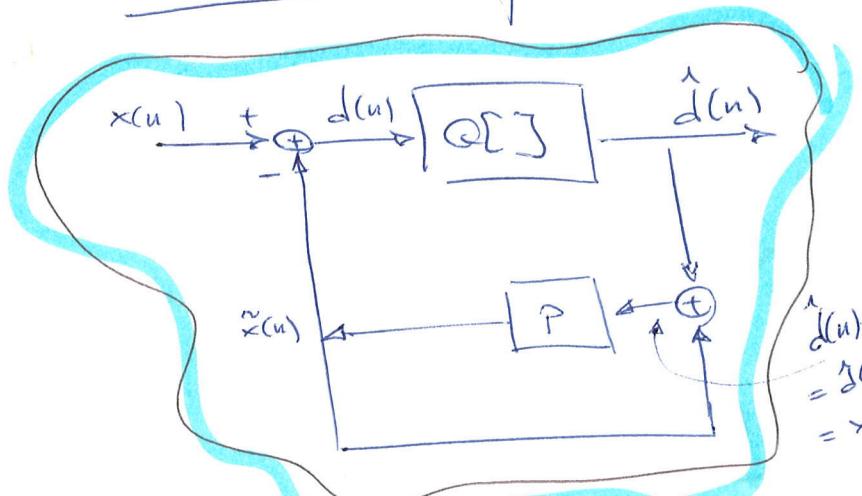
$$\text{El error es: } (S/N)_D = \cancel{\frac{\sigma_x^2}{\sigma_e^2}} \quad \Rightarrow \quad (S/N)_D = \frac{\sigma_x^2}{\sigma_e^2} = \frac{\sigma_x^2}{\sigma_e^2} \quad \frac{\sigma_e^2}{\sigma_{el}^2}$$

$$(S/N)_D = G_p (S/N)_Q \frac{\sigma_e^2}{\sigma_{el}^2}$$

Buen predictor  $\Rightarrow$  el error tiende a ser ruido blanco  $\Rightarrow$  el error n.º de blanco ( $u \uparrow$ )

$$\text{Filtros} \rightarrow \text{misma señal} \Rightarrow \frac{\sigma_e^2}{\sigma_{el}^2} = \frac{1}{G_p} \Rightarrow (S/N)_D = (S/N)_Q \quad \square$$

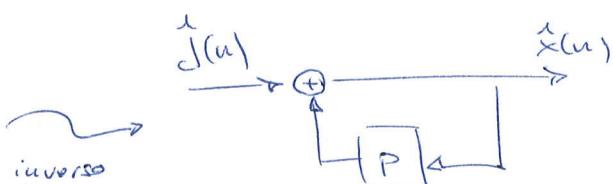
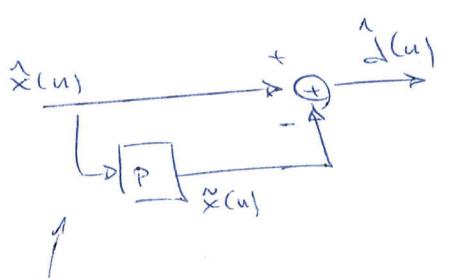
Cambiar de ecuaciones: (\*3.12\*)



$$\begin{aligned} d(u) + \hat{x}(u) &= x(u) \\ &= d(u) + x(u) - d(u) \\ &= x(u) - e(u) \end{aligned}$$

$$\hat{x}(u) = d(u) - e(u)$$

$$\hat{x}(u) = \tilde{x}(u) + \hat{d}(u) \Rightarrow \hat{x}(u) = \hat{x}(u) - \tilde{x}(u)$$

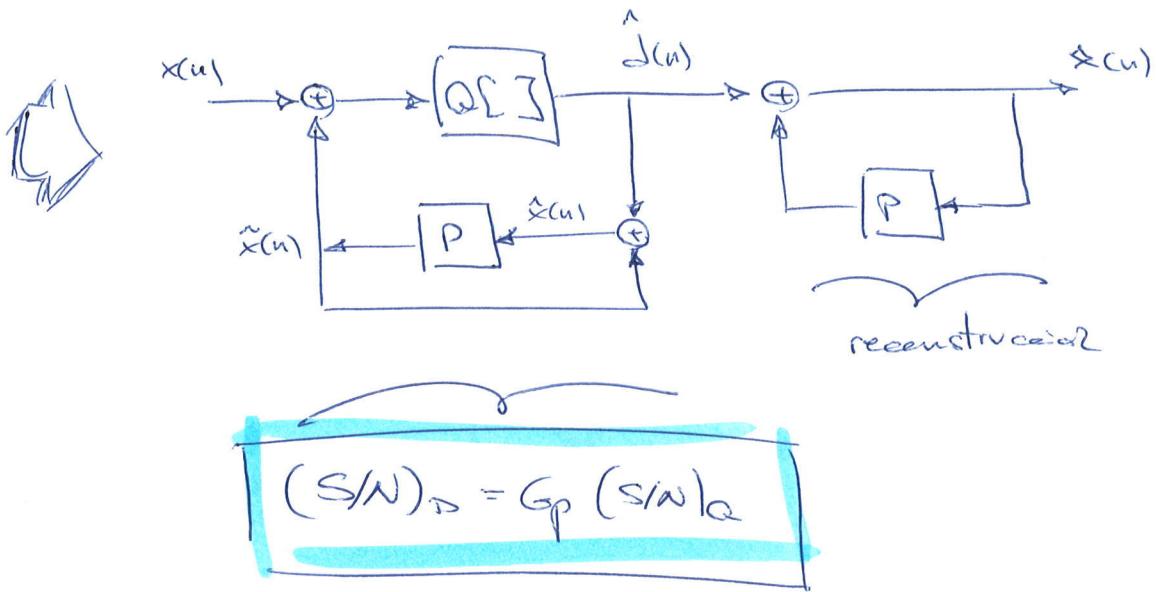


en este sistema se aplica alguno de los métodos para resolver el problema de predecir  $\Rightarrow \hat{x}_p = -\underline{T}_p^{-1} \cdot \underline{Y}_p$

$$\underline{T}_p = \begin{bmatrix} Y_x(0) & \dots & Y_x(p-1) \\ Y_x(p-1) & \dots & Y_x(0) \end{bmatrix}$$

$$\underline{Y}_p = \begin{bmatrix} Y_x(1) \\ \vdots \\ Y_x(p) \end{bmatrix}$$

$$\underline{a}_p = \begin{bmatrix} a_p(1) \\ \vdots \\ a_p(p) \end{bmatrix}$$



Hipótesis:

1)  $x(u)$  incorreto de  $e(u)$

2)  $e(u)$  blanca:

$$\begin{cases} Y_x(0) = Y_x(0) + \sigma_e^2 \\ Y_x(k) = Y_x(k) \quad k \neq 0 \end{cases}$$

Error de predicción:  $\hat{d}(u)$

$$\Rightarrow \sigma_d^2 = \sum_{k=0}^P \alpha_p(k) Y_x(k) = Y_x(0) + \sigma_e^2 + \sum_{k=1}^P \alpha_p(k) Y_x(k)$$

$Q\{\}$   $\Rightarrow u \uparrow \Rightarrow e(u)$  incorreto de  $d(u)$

$$\hat{d}(u) \xrightarrow{Q\{\}} \hat{d}(u)$$

$$\Rightarrow \sigma_d^2 = \sigma_d^2 + \sigma_e^2 \Rightarrow \sigma_d^2 = Y_x(0) + \sum_{k=1}^P \alpha_p(k) Y_x(k) \quad \left| \begin{array}{l} \text{potencia del} \\ \text{error de} \\ \text{predicción} \end{array} \right.$$

$$G_p = \frac{\sigma_x^2}{\sigma_d^2} = \frac{Y_x(0)}{Y_x(0) + \sum_{k=1}^P \alpha_p(k) Y_x(k)}$$

Se puede expresar en función de los coeficientes de correlación:

$$P_x(k) = \frac{\gamma_x(k)}{\gamma_x(0)} \quad |P_x(k)| \leq 1$$

$$\Rightarrow G_p = \frac{l}{l + \sum_{k=1}^P a_p(k) P_x(k)} \quad \text{ganancia óptima}$$

Senal blanca  $\Rightarrow P_x(k) = 0 \Rightarrow G_p = l$

Ejemplo:  $P=1$

$$a_p(l) = -\frac{\gamma_x(l)}{\gamma_x(0)} = -\frac{\gamma_x(l)}{\gamma_x(0) + \sigma_e^2} = -\frac{P_x(l)}{1 + \left(\frac{\sigma_e^2}{\gamma_x}\right)} \approx \frac{(S/N)_D}{1 + \frac{1}{(S/N)_D}}$$

$(S/N)_D$ 通常mente muy alta ( $> 30dB$ )  $\Rightarrow a_p(l) = -P_x(l)$

$$\Rightarrow G_p \approx \frac{1}{1 - P_x^2(l)} \quad \left( P_x^2(l) \rightarrow 1 \Rightarrow G_p \rightarrow \infty \right)$$

Voz no estacionaria  $\Rightarrow$  la covariación diferencial se tiene en cuenta la variación temporal de la autocorrelación

a) Autocorrelación media de la voz

Máx:  $G_p < 9 dB$  (prácticas:  $\approx 6 dB$ )

Hacer un promedio temporal de los estimadores  
⇒ predictor fijo

(\* 3.17\*)

### b) Análisis localizado (autocorrelación localizada)

Reavivar un predictor adaptativo:

⇒ ↗ . ↑ complejidad  
    ↗ . ↑ G

$G_p < 24 \text{ dB}$  (precisiones:  $\sim 15 \text{ dB} \rightarrow 2,5 \text{ bits}$ )

## 2.9.- CUANTIFICACIÓN DIFERENCIAL ADAPTATIVA (ADPCM)

- } • predictor adaptativo  
} • bucle de error adaptativo

(Bib: Rechner 226-232)

(ADPCM = Adaptive Differential Pulse Code Modulation)

Tecnología:

• Cuantificación adaptativa  $(A(n), G(n))$ :

- AQB: directa

- AQF: retrasada (backwards)

• Predictor adaptativo: (actualiza coeficientes cada 10-20 ms)

- APF: directa

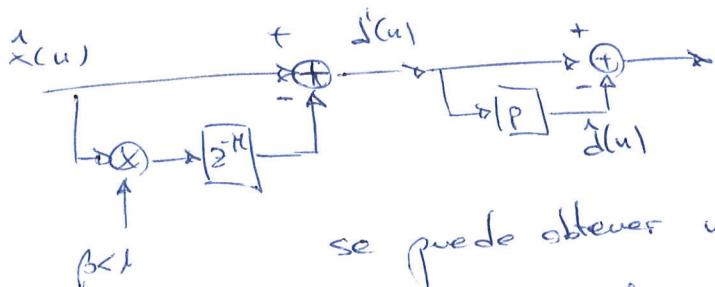
- APB: retrasada

En telefonía, ADPCM → G.721 (ITU-T)

$$\left. \begin{array}{l} f_m = 8 \text{ kHz} \\ R = 32 \text{ kbps} \\ \text{AQB, APB} \end{array} \right\} (* 3.23-3.25*)$$

Esquemas diferenciales aplicados a la señal de voz (predicción adaptativa):

Al aplicarlos a frames sonoros, se puede obtener una mejor codificación a cuenta que los frames sonoros son periódicos



$$N = \text{periodo del fundamental} (>p)$$

se puede obtener un  $\beta$  óptimo que nos dé  $\hat{x}(n)$  a potencia mínima

La mejora es independiente de lo obtenida con el predictor

$\mu \rightarrow$  fundamental

$p \rightarrow$  formantes

("predictor de 2 etapas")

## 2.19.- MODULACIÓN DELTA

Bib: Robiner 216-221

Es un tipo de modulación diferencial, el más sencillo:

- }. predictor de orden 1 ( $p=1$ )
- }. cuantificador de 1 bit ( $B=1$ )

$$\hat{x}(n) = -\alpha(1) \hat{x}(n-1) \quad (\times 3.18 - 3.29*)$$

$$\alpha(1) = -\rho_x(1) \Rightarrow G_p = \frac{1}{1 - \rho_x^2(1)}$$

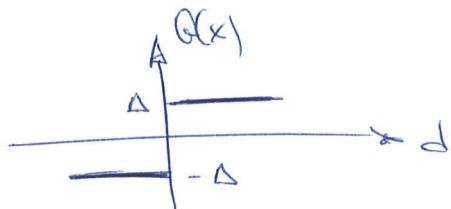
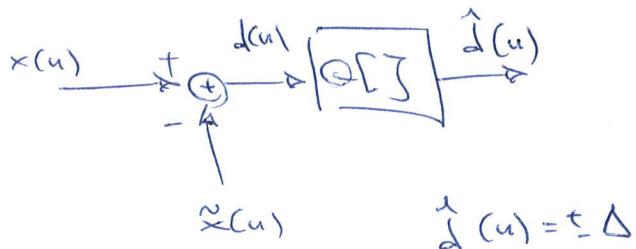
¿Aumentar  $\rho_x(1) \rightarrow 1?$   $\Rightarrow$  subir {frecuencia de muestreo  
velocidad binaria}

Simplificación:  $F = \frac{f_s}{2W}$  = factor de sobreexceso

$$x(t) \rightarrow W$$

$$\rho_x(\lambda) \rightarrow \lambda \Rightarrow \alpha(\lambda) \rightarrow -\lambda \Rightarrow \hat{x}(u) = \hat{x}(u-1)$$

$$\hat{x}(u) = \hat{x}(u) + \hat{d}(u) = \hat{x}(u-1) + \hat{d}(u)$$



$$\hat{d}(u) = \Delta \cdot \text{sign} (\hat{x}(u) - \hat{x}(u-1))$$

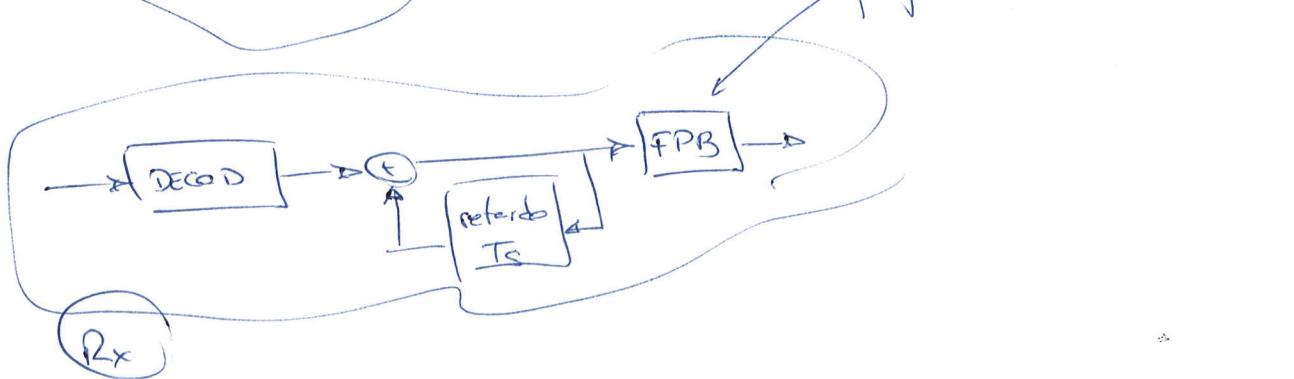
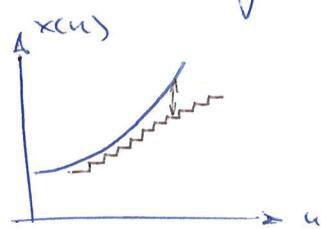
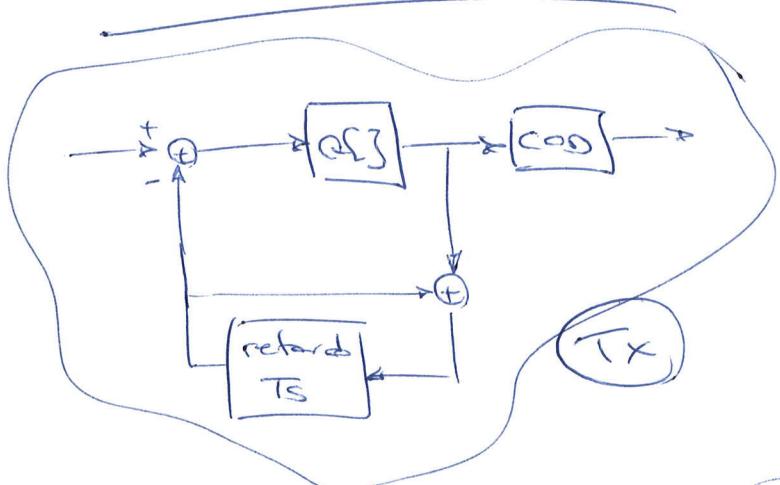
$$\Rightarrow \hat{x}(u) = \hat{x}(u-1) + \Delta \text{sign} (\hat{x}(u) - \hat{x}(u-1))$$

Si  $\max \{ |x(u) - x(u-1)| \} \leq \Delta \Rightarrow \text{erreur} \leq \Delta$

(cas de granularité)

→ variations de la signal lentes

$\max \{ |x(u) - x(u-1)| \} > \Delta \Rightarrow \text{risque de saturation}$

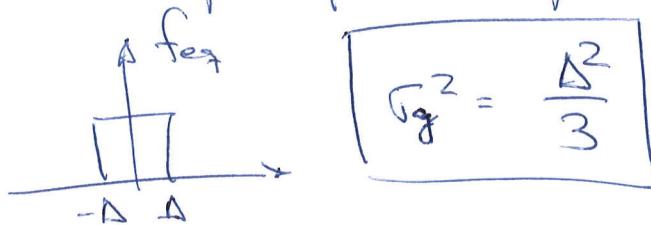


## Quantificación del análisis:

- Alora  $\Delta$  no es la separación entre niveles, sino que es  $2\Delta$
- Ruido grande no uniforme a  $[-\Delta, \Delta]$

## Simplificación:

- Se supone que el ruido grande es uniforme en  $[-\Delta, \Delta]$



Ruido de sobreexig.: perímetro

$$\sigma_s^2 = k \cdot S \cdot e^{-S/2}$$

(\* 3.29\*)

$$S = \frac{\Delta}{\sigma_g}$$

$$k = \frac{243}{4\sqrt{2\pi}} \frac{D_1^2}{D_2} \quad \text{Jouant-Noll:} \quad k = \frac{243}{4} \sqrt{\frac{\pi}{2}} \frac{D_1^2}{D_2}$$

Factor de mejora en el nivel de ruido  $\rightarrow$  filtrado peso bajo posterior

$\hat{x}(n) \rightarrow G_e^2$  (error de cuantificación)

$$\text{F.P.B} \rightarrow \hat{x}_{LPF}(n) \rightarrow \sigma_r^2 = \frac{1}{F} G_e^2$$

$G_e$  AWGN

## 2.11.- MODULACIÓN DELTA ADAPTATIVA

Bib: Robins 221-224 (\* 3.21 - 3.22\*)

- ⇒
  - . Jouant
  - . estructuras redimensionables

- Si key saturación: se mantiene durante muchas muestras el signo de la diferencia ( $\pm \Delta$  p.ej.)

$$\Rightarrow \boxed{\Delta(n) = \lambda \cdot \Delta(n-1)}$$

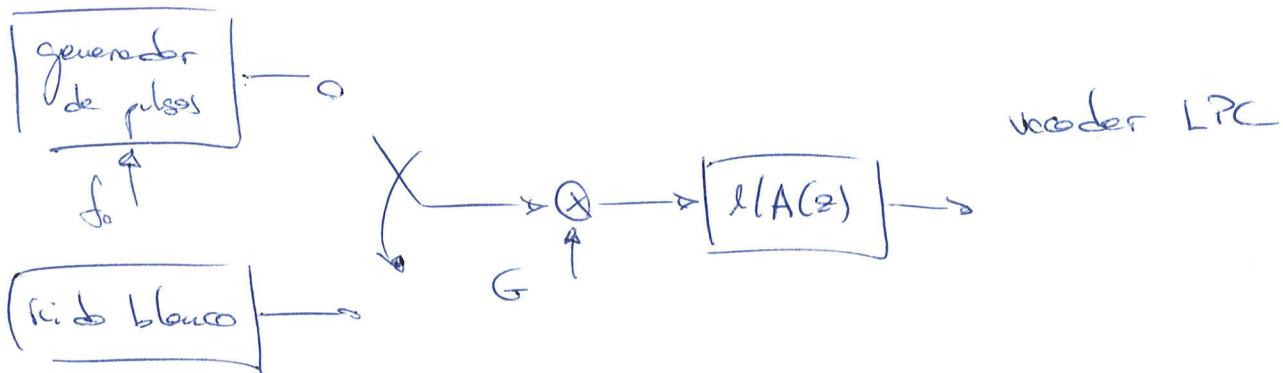
- En ruido grande hay alteración de signo

$$\mu = \begin{cases} P > 1 & c(u) = c(u-1) \\ Q < 1 & c(u) \neq c(u-1) \end{cases}$$

Estudio de estabilidad  $\Rightarrow P \cdot Q \leq 1$

### 3.- Vocoders Mixtos o Hibridos

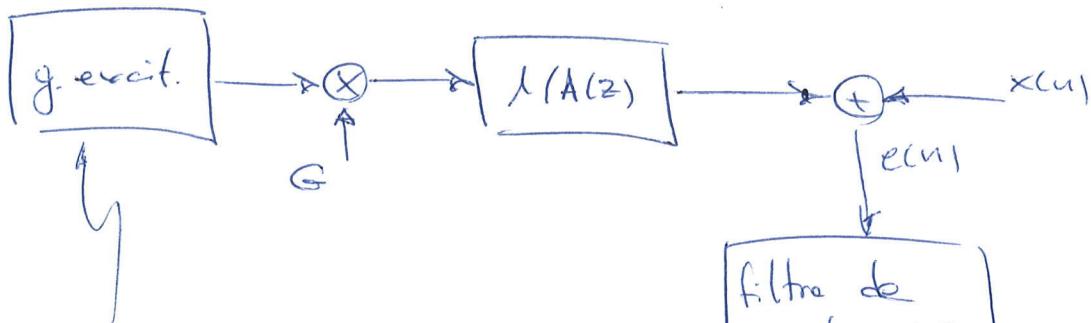
- Vocoder: sistema de cuantificación paramétrica. Se cuantifican los parámetros de un modelo.



- dan una señal muy poco natural, muy sintética
- buscar compromiso entre codificaciones de forma de onda y paramétricas

#### ANÁLISIS POR SÍNTESIS:

Para cada frame de señal se produce una señal que se compara con la señal de voz que se quiere cuantificar



habitual: librería de señales excitadoras



se prueba a todos y se queda con el que menor error

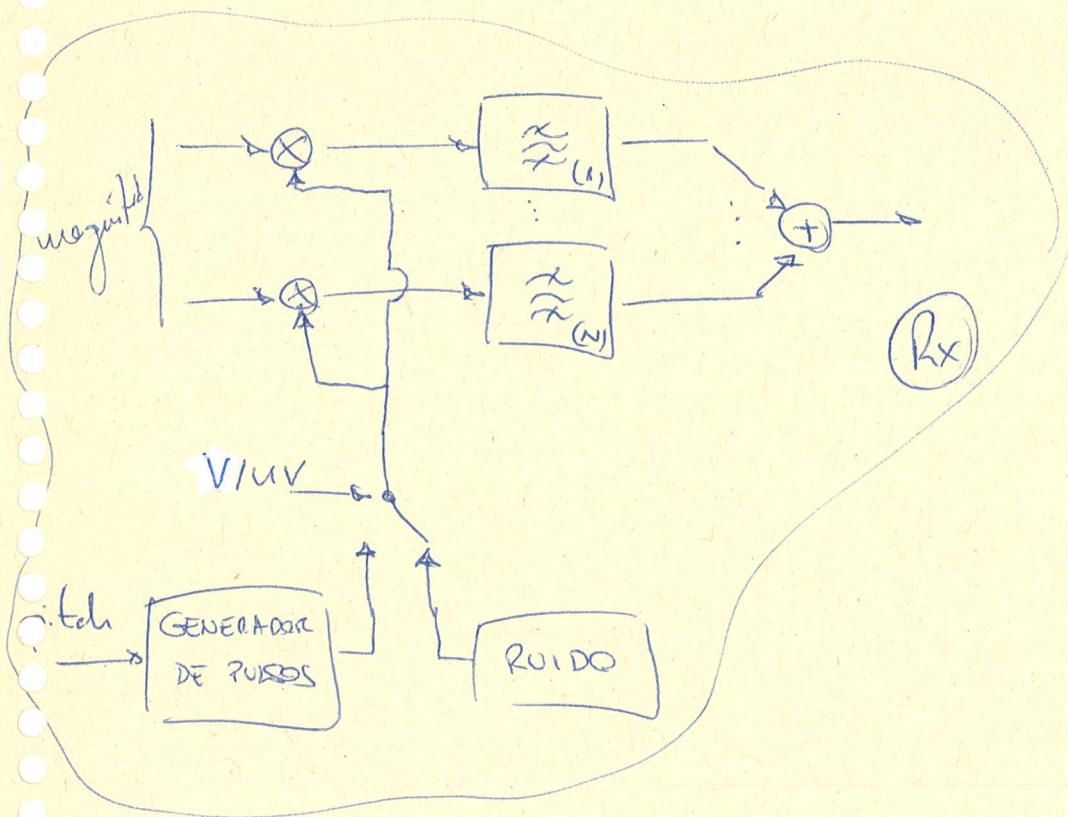
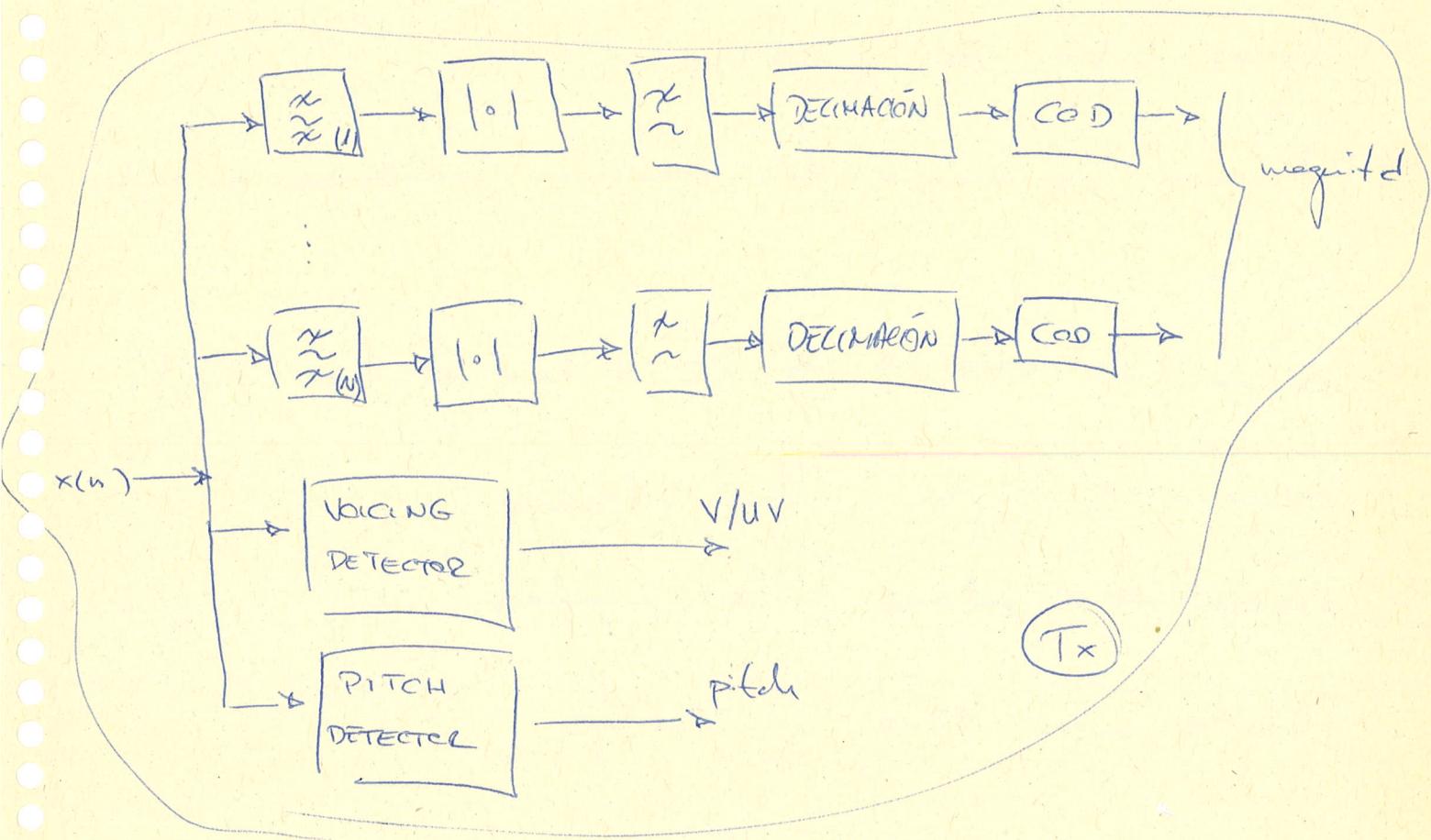
$\Rightarrow$  se envía el índice de la señal excitadora escogida

✓ esto permite recuperar el carácter

$\Rightarrow \text{GSTH}$



# VOCODER DE CANAL

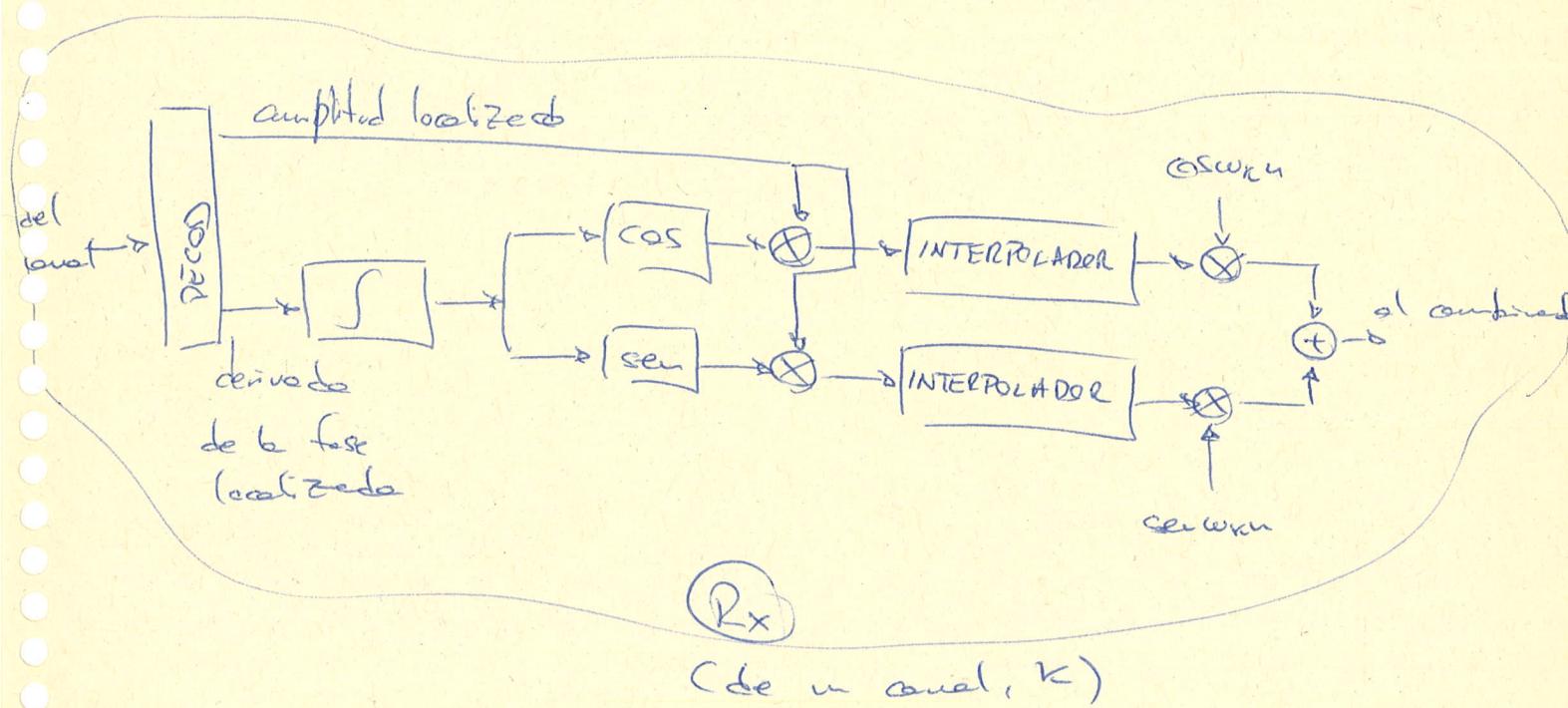
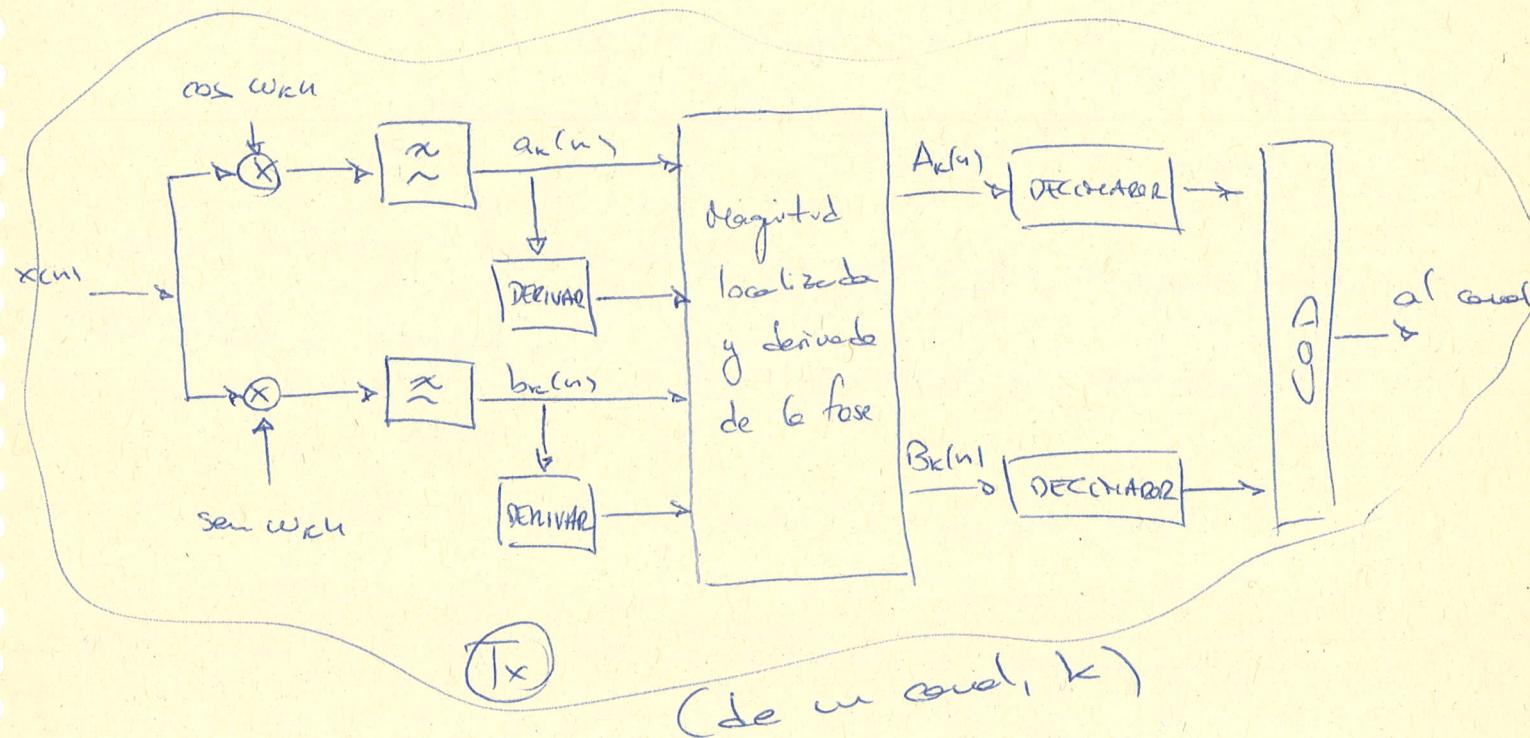


- Banda de filtros para banda ancha de banda típicos entre 100 y 300 Hz
- Estimaron las magnitudes espectrales localizadas en cada banda.
- Banda 0-4 kHz  $\Rightarrow$  16-29 filtros de fase lineal
- Frecuencias bajas: filtros estrechos  
frecuencias altas: filtros anchos
- Bits/bytes entre 1200 y 3200 bps

## VOCODER DE FASE

- Similar al vocoder de canal, pero no estima el pitch, sino la derivada de la fase o la salida de cada filtro del banco.

Filtro del banco.



## Problemas del vector de fase:

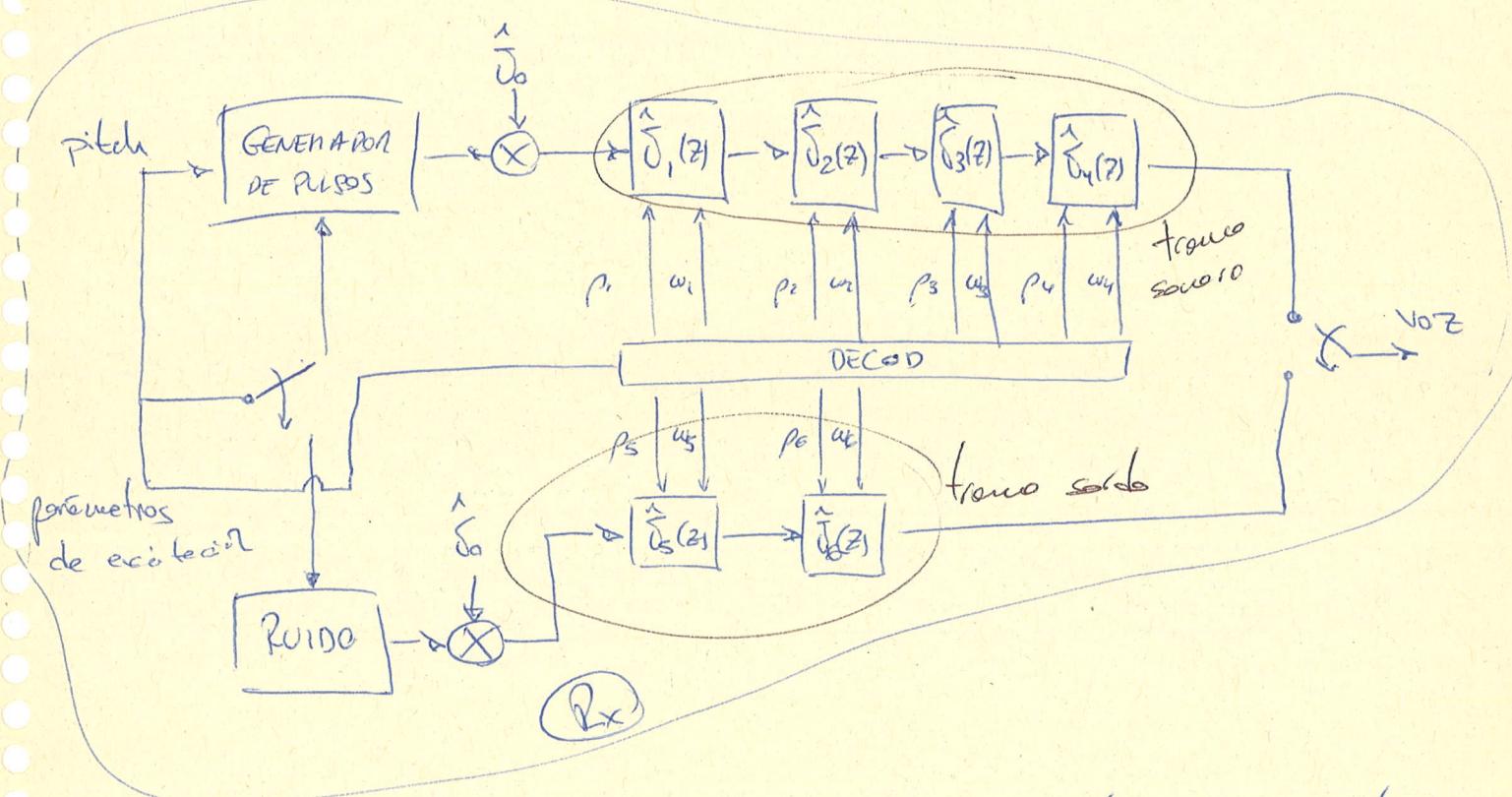
- Se ha supuesto que sólo pasa una sinusode a una sola frecuencia por cada filtro del banco.
- Esto puede provocar una mala estimación si la respuesta del filtro no es plana.
- La estimación también puede fallar si una sinusode cae entre los filtros adyacentes.
- Integración numérica en el cálculador produce errores y la fase inicial no se conoce  $\Rightarrow$  "dispersión" o desvío de fase.
  - ~ Los pérdidas de coherencia en la fase.
- Hay mejoras al vector de fase que eliminan estos problemas.

## Vocoder de formantes

- Tipo de vocoder de canal que estima los 3 o 4 primeros formantes de un segmento y sus anchos de banda.
- Se transmite esa información junto al pitch.
- Caracterización de cada formante en un tramo de voz.

$$\hat{J}_k(z) = \frac{J_k}{(1 - p_k e^{i\omega_k z^{-1}})(1 - p_k e^{i\omega_k z^{-1}})} \quad k=1, 4, \dots$$

$\omega_k = 2\pi f_k T$



- Los formantes se estiman por predicción lineal o análisis espectral.

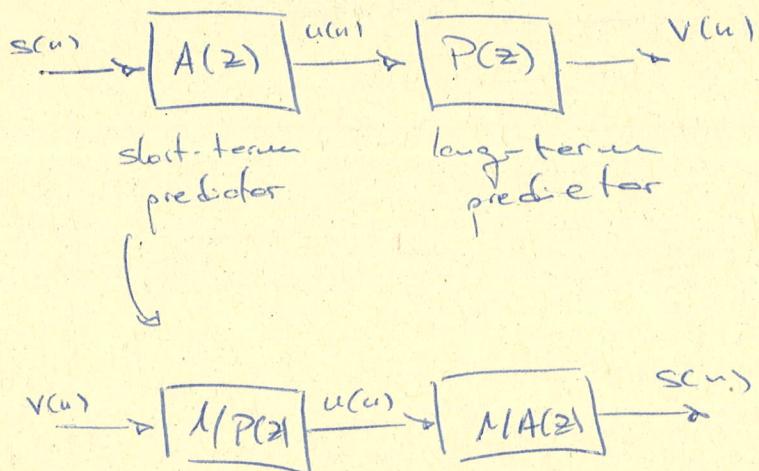


## -MPC (Multi-Pulse Coding)

- Considera la mezcla de sonido sordo y severo.
  - El pitch produce correlación de periodo largo que se dice con el filtro LPC
- segundo predictor lineal (filtro de pitch):

$$P(z) = 1 - \sum_i \beta_i z^{-N-i} \quad N = \text{periodo de pitch}$$

- Buena calidad → 9,6 kbps
  - Impulsos: posición y amplitudes determinadas a través de sorteo
- patrones de pulsos equiespaciados → RPE



## - MELP (Mixed Excitation LPC)

- Excitación multibanda: representar la voz como suma de sinusoides armónicas
- Usar diversos tipos de pulsos y sumar para los 8 formantes (4-10) bandas de frecuencia, que pasan por filtros variables con el tiempo y se suman para formar la señal de banda completa.
- Impulsos apagadores debidos al jitter del pitch, voz "crijiente" y voz ligátrica
- Ancho de banda de resonancia variable con estimación del periodo de pitch
- Tramas de 22,5 ms  
→ 54 bits/trama ⇒ 2,4 kbps

## VOCODERS

- RELP (Regular Excited Linear Predictor)

RPE-LTP (Regular Pulse Excitation - Long Term Prediction)

- GSM

- Segmentos: 20 ms o 506 pos., 160 muestras

- LPC,  $\gamma = 8$

- Log-Arc:  $LAR(i) = \log \frac{1+k_i}{1-k_i}$

- Filtro de predicción de referido largo: evalúa 4 veces por segmento  $\rightarrow$  5 ms (40 muestras)

- Pitch  $\in [40, 120] \rightarrow$  7 bits

- Gancor  $\rightarrow$  2 bits

- GSM  $\rightarrow$  13 kbps

## - CELP (Code Excited Linear Predictor)

- Análisis por autocorrelación, predictor lineal y cuantificación vectorial con pesos
- Periodo corto: LPC-10
- Periodo largo: VQ o diccionario adaptativo
- Error de predictor lineal y pitch VQ se cuantifican en diccionario de secuencias estocásticas (múltiples de WGN)
- Se transmite el índice del diccionario y la ganancia
- Filtro perceptual atenúa las frecuencias inaudibles y amplifica frecuencias a las que el oído es más perceptivo.
- Frames de 30 ms (240 muestras)
- 4,8 kbps  $\Rightarrow$  144 bits / frame

## -VSELP (Vector Sum Excited Linear Prediction Coding)

- 3 fuentes de excitación:

→ 2 licencias

$$\Rightarrow 8 \text{ kbps} = 2 \text{ licencias de } 128 \text{ vectores}$$

$$\Rightarrow 4,8 \text{ kbps} = 1 \text{ licencia de } 2048 \text{ vectores}$$

→ licencia adaptativa a filtro de periodo largo

- Filtro de síntesis: LPC-10 → cuantificación escalar

- Tramas de 20ms

- Excitación actualizada cada 5 ms

- 8 kbps  $\Rightarrow$  40 muestras por trama

- Gama de las fuentes de excitación: VQ

