

RECOMENDACIÓN UIT-R F.1112-1*

TRANSMISIÓN DE SEÑALES VOCALES DIGITALIZADAS EN SISTEMAS QUE FUNCIONAN POR DEBAJO DE UNOS 30 MHz

(Cuestión UIT-R 164/9)

(1994-1995)

La Asamblea de Radiocomunicaciones de la UIT,

considerando

- a) que las comunicaciones vocales en la banda de ondas decamétricas utilizan canales de 3 kHz;
- b) que en el caso de algunas comunicaciones la seguridad es esencial;
- c) que la aleatorización constituye el único medio de alcanzar un nivel de seguridad suficiente;
- d) que el nivel de seguridad exigido pueda lograrse fácilmente utilizando técnicas de digitalización de la palabra;
- e) que, en consecuencia, conviene realizar códecs de señales vocales (vocoderes) asociados a módems en la banda de ondas decamétricas;
- f) que para los canales de buena calidad en la banda de ondas decamétricas, la velocidad binaria admisible es como máximo de 4 800 bit/s;
- g) que la interferencia y los efectos de propagación tales como el desvanecimiento, suponen para las comunicaciones digitales un incremento de la proporción de bits erróneos, lo que exige la utilización de procedimientos de corrección (códigos correctores de errores-entrelazado),

recomienda

- 1** que para las comunicaciones de corto alcance (que utilizan ondas de superficie), se empleen vocoderes de velocidad 2 400 bit/s o 4 800 bit/s;
- 2** que para las comunicaciones a larga distancia (que utilizan ondas ionosféricas) se empleen, según la calidad del enlace, vocoderes de velocidad 2 400 bit/s o vocoderes de velocidad 600-800 ó 1 200 bit/s con codificación para corrección de errores;
- 3** que los sistemas de radiotelefonía digital utilizados se ajusten a las especificaciones generales señaladas en el Anexo 1 y a las especificaciones particulares indicadas en los Anexos 2, 3, y 4, según el tipo de vocoder utilizado.

ANEXO 1

Consideraciones generales sobre los sistemas de radiotelefonía digital en la banda de ondas decamétricas

Un sistema de radiotelefonía digital en ondas decamétricas se compone de un circuito radiotelefónico tradicional, de un codificador de señales vocales (vocoder), de un dispositivo opcional de aleatorización y de un módem en ondas decamétricas. En la Fig. 1 se representa el esquema básico de este sistema.

La señal vocal transmitida llega al codificador vocal donde se analiza y transforma en flujo digital. Este flujo puede aleatorizarse y llega a continuación a un módem que le da la forma adecuada para su transmisión en la banda de frecuencias del canal telefónico. El flujo digital procedente del receptor del módem se decodifica, si es necesario, y pasa al decodificador vocal donde se reconstruye la señal vocal.

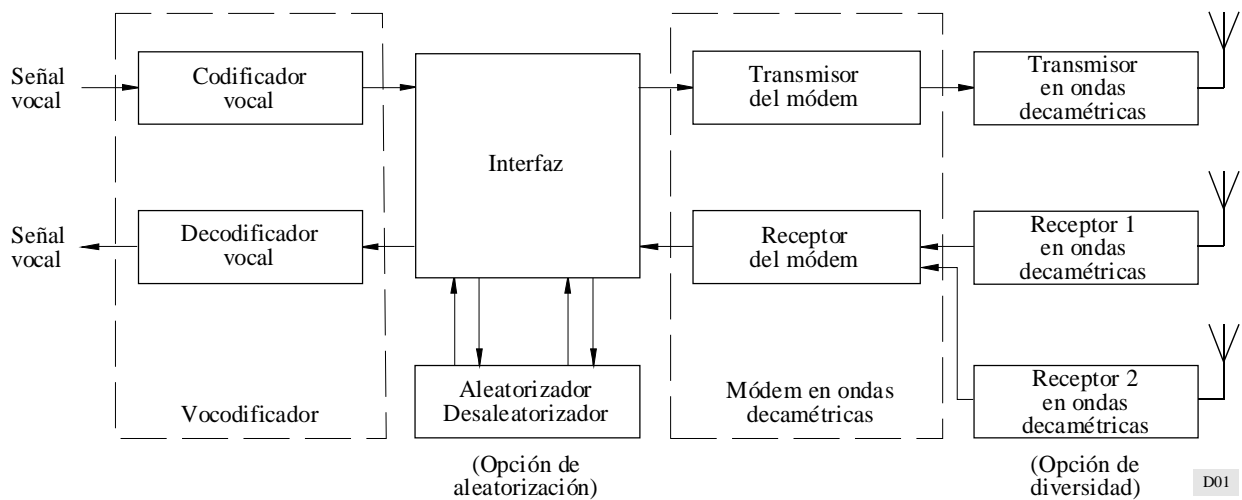
* Esta Recomendación debe señalarse a la atención de la Comisión de Estudio 8 de Radiocomunicaciones.

En el caso de proporciones de errores elevadas a las cuales no pueden hacer frente los vocodificadores de velocidad binaria reducida (2 400 ó 4 800 bit/s), es preciso utilizar vocodificadores de velocidad binaria *muy* baja (600-800 ó 1 200 bit/s) con codificación para corrección de errores.

Por otra parte, es posible utilizar recepción con diversidad que se efectúa con ayuda de dos receptores conectados a antenas con diversidad espacial, dando un tratamiento posterior a la señal en la parte receptora del módem.

Existen numerosos tipos de vocodificador con velocidad $\leq 4\,800$ bit/s, en particular los vocodificadores de canales, los vocodificadores ortogonales y los vocodificadores de predicción lineal; en los Anexos 2, 3 y 4 se describen sistemas que utilizan estos tipos de vocodificadores.

FIGURA 1
Esquema general de un radioteléfono digital en ondas decamétricas



Los tres tipos de vocodificadores tienen una velocidad binaria deseada de 2 400 bit/s y una inteligibilidad de sonidos más del 90% en buenas condiciones de transmisión. En términos de calidad son aproximadamente equivalentes para dicha velocidad binaria.

Todos los vocodificadores se presentan con los módems asociados, que en los tres casos se trata del tipo «módem paralelo» (varias subportadoras con modulación independiente en la banda de audio).

En el vocodificador de canal y el vocodificador ortogonal, el breve periodo del símbolo elemental transmitido por el módem (13,3 ms y 8,33 ms, respectivamente) los hace menos resistentes a las severas condiciones de propagación de la banda de ondas decamétricas, caracterizadas por la aparición de trayectos múltiples con retardos de unos pocos milisegundos, lo que provocará una importante interferencia entre símbolos.

En el vocodificador LPC 10, donde el módem asociado es el descrito por STANAG 4197, el periodo de símbolo es 22,5 ms. En consecuencia, este módem resulta menos afectado en tales condiciones.

Además, en el vocodificador LPC 10 algunos bits de trama importantes son protegidos mediante un código de corrección poderoso (código Golay) lo que da lugar a que la degradación de la calidad de funcionamiento a medida que aumenta la proporción de bits erróneos (BER) es más gradual que con los anteriores vocodificadores: la inteligibilidad de sonidos se reduce hasta el 80% para una BER del 2%, en comparación con el 1% para otros vocodificadores.

Debe garantizarse un funcionamiento satisfactorio en malas condiciones. Ello se logra fácilmente con el vocodificador LPC 10 utilizando simultáneamente:

- comprensión de datos para reducir la velocidad binaria deseada a 600/800/1 200 bit/s (lo que supone una disminución en la calidad intrínseca del vocodificador);
- inserción de redundancia, lo que aumenta la resistencia frente a los errores.

Con esta disposición, en la versión de 800 bit/s, no se alcanza el umbral de inteligibilidad de sonidos del 80% hasta que la BER rebasa el valor del 5%.

ANEXO 2

Sistema de radiotelefonía digital que utiliza un vocodificador de canales

1 Descripción del sistema

Este sistema está constituido por un circuito radiotelefónico convencional por ondas decamétricas y un terminal vocal digital.

La señal de conversación de entrada procedente del micrófono se analiza y digitaliza en el vocodificador de canales. A continuación, la señal digital se aplica, a través del interfaz, al aleatorizador donde se mezcla con las señales de aleatorización. La salida del aleatorizador se encamina a través del interfaz al modulador, donde se convierte en una señal de frecuencias vocales mediante el método de modulación MDF-MDP-4D.

La salida del receptor es una señal de audio, la cual se aplica al demodulador que la convierte en códigos digitales. La señal de código digital se dirige, a través de la interfaz, al desaleatorizador, donde se convierte en los códigos digitales originales. Esta señal digital se aplica a través del interfaz, al vocodificador de canales, que reconstruye la señal telefónica, que se entrega seguidamente al receptor telefónico.

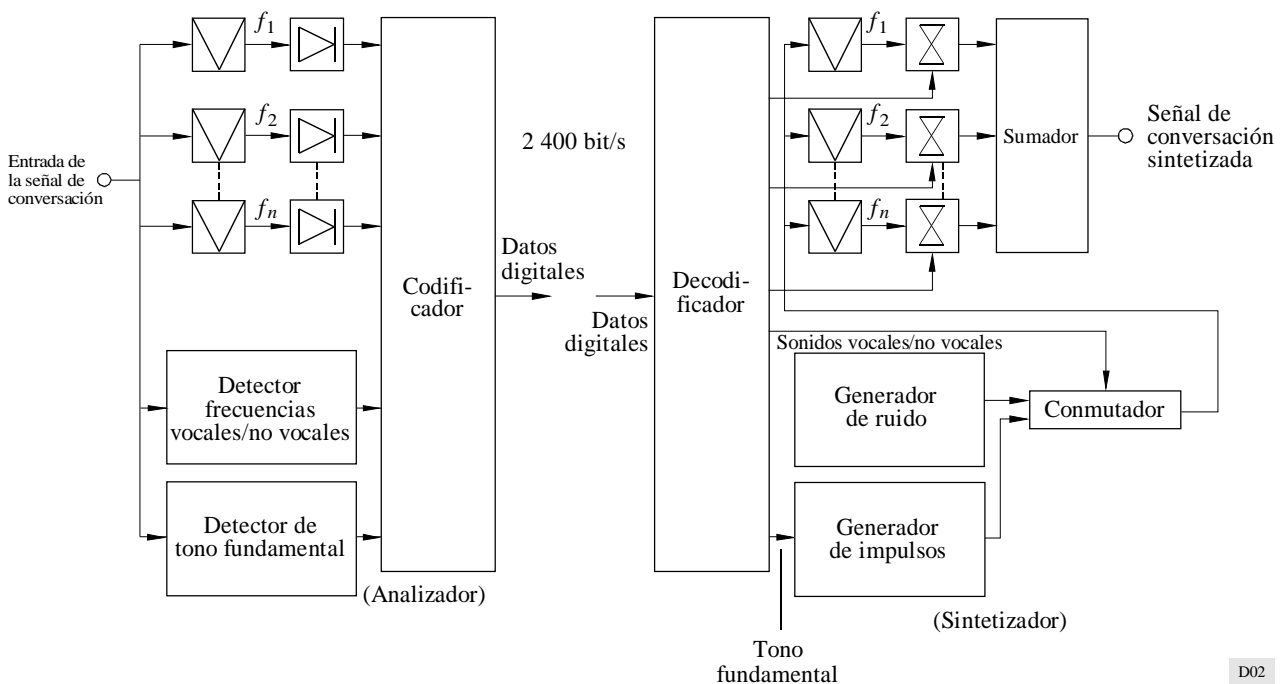
2 Vocodificador de canales

2.1 Teoría

El vocodificador de canales divide la señal de frecuencias vocales comprendidas entre 300 y 3000 Hz, aproximadamente, en varias bandas espectrales contiguas y mide la energía en cada banda. Estas mediciones se codifican y se transmiten.

En la Fig. 2, se representa el diagrama de bloques teórico del vocodificador de canales. En la sección de análisis, se utilizan diversos filtros paso banda (FPB – procesado de filtro equivalente) para separar las sub-bandas de conversación y extraer la componente de frecuencia de cada banda espectral. Se miden las salidas de los FPB para determinar el nivel de cada banda. Al mismo tiempo, se detectan los sonidos sonoros y sordos y el tono fundamental. En el analizador se muestrean y cuantifican estos parámetros, obteniéndose una señal codificada de 2 400 bit/s para la transmisión.

FIGURA 2
Diagrama de bloques del vocodificador de canales



En la sección de síntesis se utilizan, un generador de ruido y un generador de impulsos para excitar el sintetizador del espectro. El generador de ruido actúa sobre los sonidos sordos y el generador de impulsos sobre los sonidos sonoros. Se ajusta la frecuencia de salida del generador de impulsos a un valor casi igual al del tono fundamental. Se analizan las salidas del generador de ruido o del generador de impulsos mediante una combinación de filtros paso banda, similar a la utilizada en la sección de transmisión. Se multiplican y suman los niveles de las bandas espectrales analizadas para recuperar la señal de conversación.

2.2 Realización

El terminal vocal digital utiliza la FFT (transformada rápida de Fourier) para realizar el análisis espectral de la señal telefónica. El espectro calculado se subdivide en varias bandas espectrales cuya anchura es equivalente a la de cada filtro paso banda. Se promedian las bandas espectrales o canales para determinar el nivel de cada una de ellas. Se detecta la frecuencia obteniendo la autocorrelación máxima y se obtienen los sonidos voz/no voz basándose en el nivel de los valores máximos de la función. Por otro lado, se sintetiza la señal telefónica generando la respuesta impulsiva de los filtros paso banda mediante un filtro digital RIF (respuesta al impulso finita). El resultado se multiplica por el nivel de salida de cada banda espectral. Por último, se suman las ondas de todas las bandas espectrales para obtener la señal de conversación.

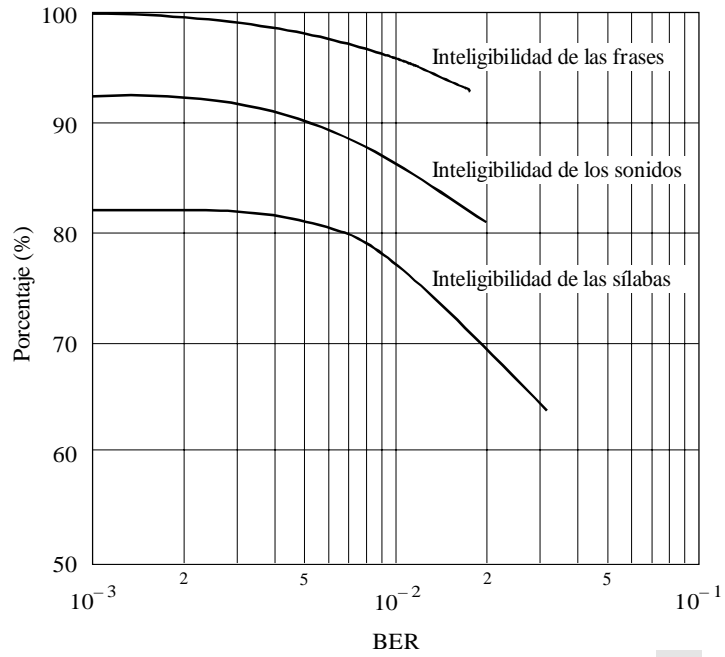
3 Módem MDF-MDP-4D

El módem MDF-MDP-4D empleado en este sistema, es básicamente el mismo que se describe en la Recomendación UIT-R F.763 (véase el Anexo 1). Sus características más importantes son: velocidad de datos 2400 bit/s, utilización de 18 tonos de los cuales 16, separados entre sí 110 Hz y situados en la banda 935-2585 Hz, se modulan con MDP-4-CD (modulación por desplazamiento de fase cuaternaria-codificación diferencial), con una velocidad de señalización de 75 Bd. Se utiliza el tono de 605 Hz para la corrección de los errores de frecuencia de extremo a extremo. Se ha elegido para la sincronización el tono de 825 Hz para evitar una pérdida excesiva en el borde de la banda. Se introducen tiempos de guarda entre las tramas para combatir la propagación multitrayecto y la distorsión de tiempo de propagación de grupo.

4 Resultado de las pruebas

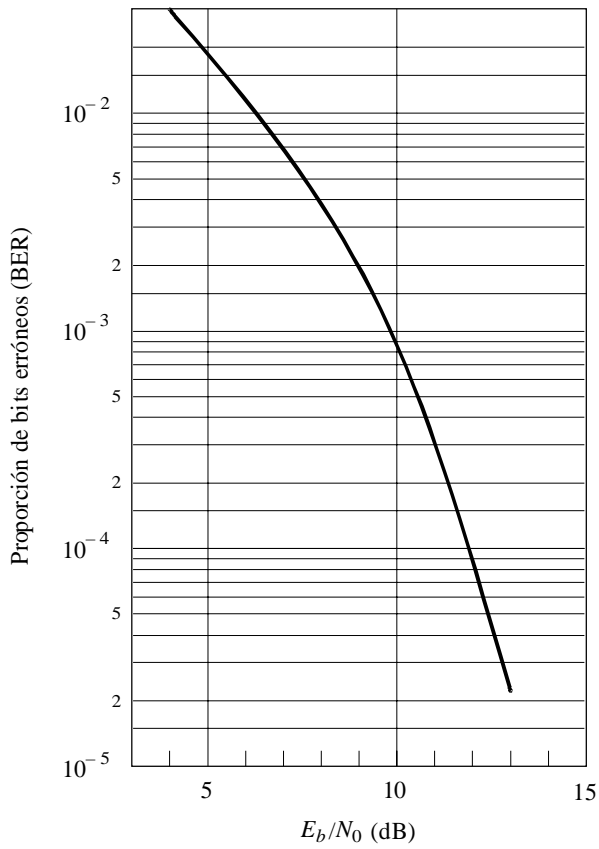
En la Fig. 3, se presentan las características de un vocodificador de canales, en términos de la inteligibilidad de las frases, los sonidos y sílabas en función de la BER. Las Figs. 4 y 5 muestran las características del módem MDF-MDP-4D en forma de BER, en función de la relación E_b/N_0 . La Fig. 4, se refiere a las características estáticas y la Fig. 5 a las características en condiciones de desvanecimiento.

FIGURA 3
Características de un vocodificador de canales



D03

FIGURA 4
Características estáticas del módem MDF-MDP-4D

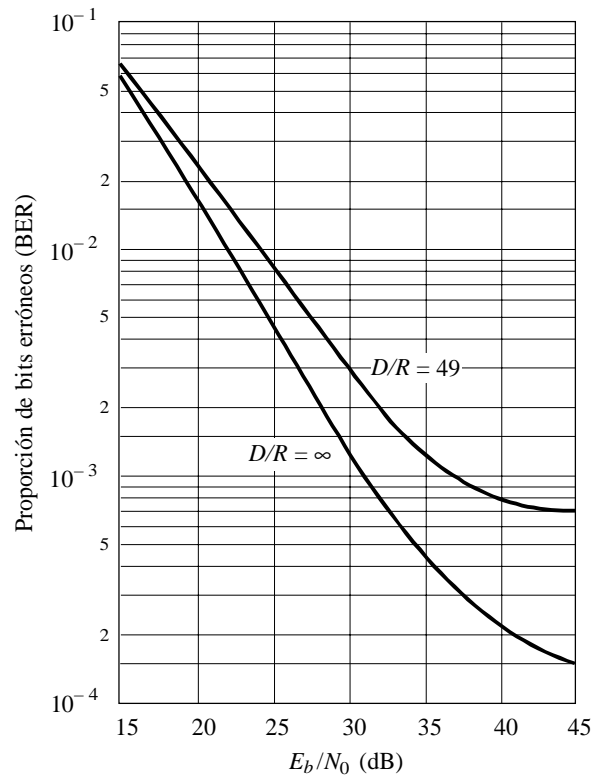


E_b : Energía por bit

N_0 : Densidad espectral de potencia de ruido

D04

FIGURA 5
Características del módem MDF-MDP-4D
en condiciones de desvanecimiento



Frecuencia Doppler: 0,5 Hz
Retardo temporal: 1 ms

D/R : Relación entre las potencias medias cuadráticas de la onda directa y la onda reflejada

D05

ANEXO 3

Sistema de radiotelefonía digital que utiliza un vocodificador ortogonal

1 Descripción del sistema

La Fig. 1 presenta un diagrama de bloques de un tipo de sistema radiotelefónico digital. El circuito comprende los equipos radioeléctricos convencionales de un enlace principal en ondas decamétricas: transmisores, receptores para doble recepción por diversidad de espacio, antenas y circuitos de enlace.

El equipo terminal de línea consiste en un vocodificador para la conversión de la señal vocal en un tren de bits a 2 400 ó 4 800 bit/s y un módem para comunicar con los sistemas radioeléctricos en ondas decamétricas.

2 Vocodificador

Se han realizado pruebas de funcionamiento en el circuito radiotelefónico digital utilizando dos tipos de vocodificador diseñados para el funcionamiento a 2400 y 4800 bit/s. El diagrama de bloques de los vocodificadores, ambos ortogonales, se muestra en las Figs. 6 y 7. La señal vocal inicial se aplica al analizador de espectro, que determina los valores Y_k de la envolvente del espectro de la señal vocal a diferentes frecuencias.

El vocodificador 4800 bit/s toma 30 muestras distribuidas uniformemente dentro de cada una de las tres secciones de la gama total de frecuencias de la señal vocal. Estas muestras se convierten en 16 coeficientes de γ_j de la descomposición de la envolvente del espectro en una serie ortogonal que, con una frecuencia de trama de 60 Hz, constituye una secuencia binaria a 3840 bit/s. Al efectuar el análisis del espectro, se extraen simultáneamente el periodo de frecuencia fundamental de la señal vocal, así como la característica de excitación del «ruido en la frecuencia fundamental», que se transmiten al doble de la frecuencia de trama, 120 Hz, con un código de 8 dígitos, utilizando así 960 bit/s del tren de bits total.

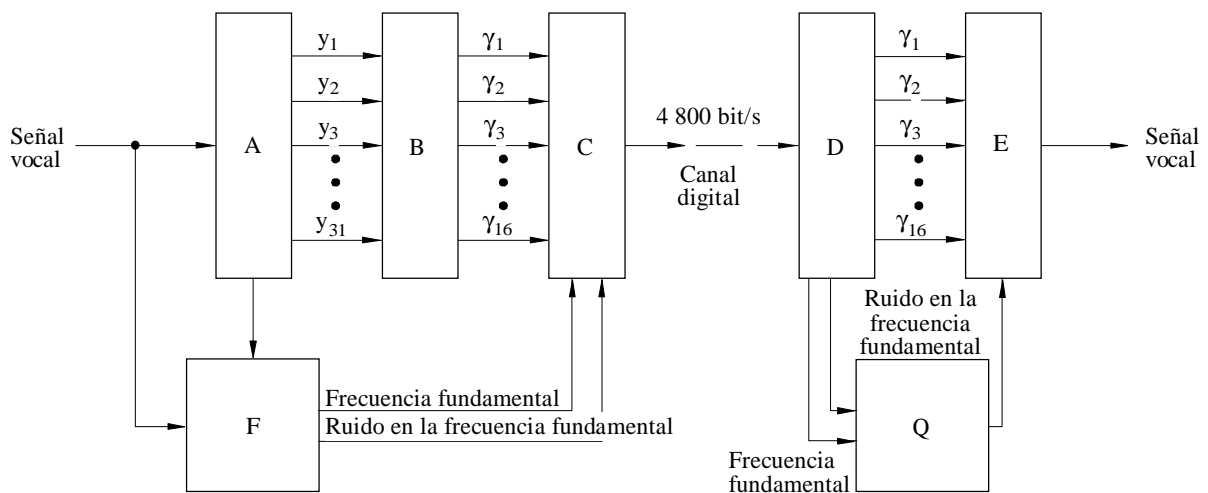
En la recepción, las señales γ_j que controlan el sintetizador de espectro ortogonal, y las señales de la frecuencia fundamental y de ruido en la frecuencia fundamental que controlan el generador de excitación del sintetizador, se extraen del tren binario recibido. El generador de excitación del sintetizador genera una secuencia de grupo de impulsos a la frecuencia fundamental que tiene un espectro uniforme, o una secuencia de impulsos pseudoaleatorios. La señal vocal se sintetiza a la salida del sintetizador.

El vocodificador 2400 bit/s (véase la Fig. 7) es ortogonal con conversión no lineal, en la cual la serie ortogonal es la raíz cuadrada de la envolvente del espectro de la voz. En este caso, el número de muestras Y_k de la envolvente del espectro es 21, y la distancia entre ellas aumenta uniformemente con la frecuencia siguiendo la curva de articulaciones iguales. Estas muestras están sujetas a la extracción de la raíz cuadrada, $\sqrt{Y_k}$, después de lo cual se determinan 10 coeficientes γ_j de la descomposición de la raíz cuadrada de la envolvente del espectro en la serie ortogonal. Los valores de estos coeficientes se transmiten cada 20 ms en el tren binario total mediante combinaciones de código de cuatro dígitos, también a intervalos de 20 ms.

En la recepción, el tren a 2400 bit/s se divide en las componentes de la señal, de las cuales las señales en la frecuencia fundamental y de ruido en la frecuencia fundamental controlan el generador de excitación del sintetizador, que es análogo al del vocodificador 4800 bit/s y, en paralelo, controlan dos sintetizadores de espectro. La salida del primer sintetizador se conecta a la salida de excitación del segundo sintetizador, y ambos elevan al cuadrado el espectro que ha de sintetizarse y establecen una relación lineal entre los espectros inicial y sintetizado de la palabra.

FIGURA 6

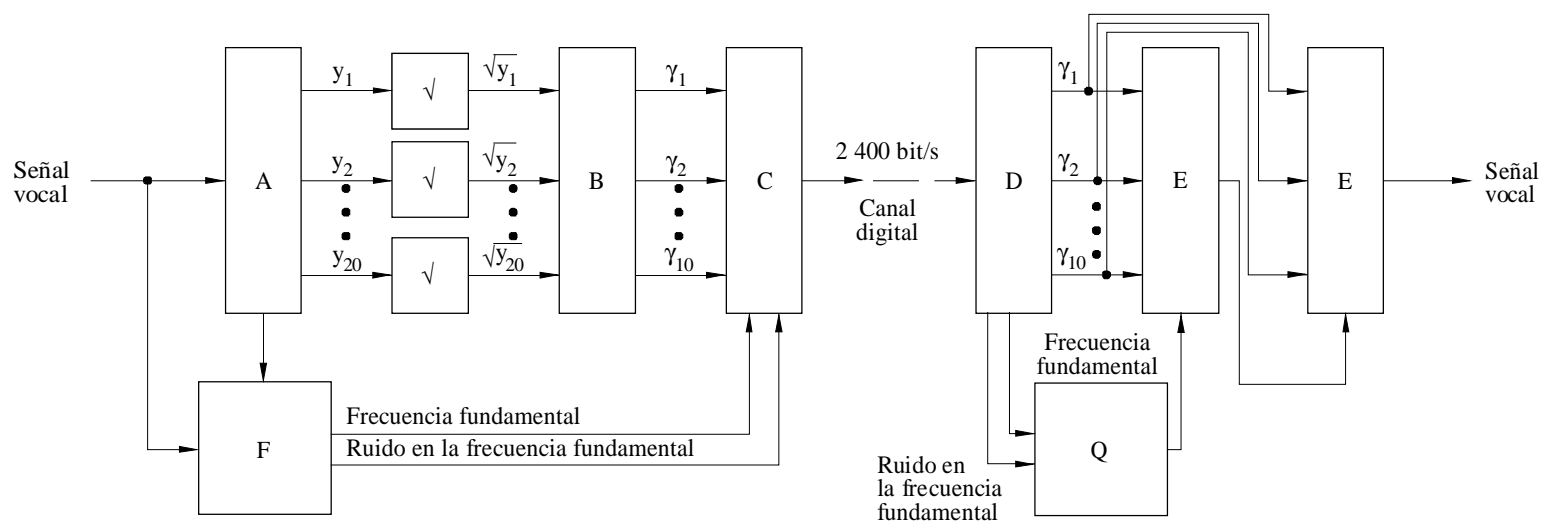
Diagrama de bloques del vocodificador ortogonal a 4800 bit/s



- | | |
|----------------------------|----------------------------------------|
| A: Analizador de espectro | E: Sintetizador |
| B: Transformador ortogonal | F: Extractor de frecuencia fundamental |
| C: Interfaz de transmisión | Q: Generador de excitación |
| D: Interfaz de recepción | |

FIGURA 7

Diagrama de bloques de un vocodificador ortogonal a 2 400 bits



- | | |
|----------------------------|----------------------------------------|
| A: Analizador de espectro | E: Sintetizador |
| B: Transformador ortogonal | F: Extractor de frecuencia fundamental |
| C: Interfaz de transmisión | Q: Generador de excitación |
| D: Interfaz de recepción | |

D07

3 Módem

El módem es un MDP-2 multicanal que utiliza señales ortogonales. Sus características básicas son las siguientes:

- velocidad de modulación de subcanal: 100 ó 120 bit/s,
- número de canales: 20,
- separación de frecuencia entre canales: 142 Hz,
- intervalo de ortogonalidad: 1/142 s,
- longitud del intervalo de protección: 1,29 ó 2,29 ms,
- método de recepción de la señal de canal: no coherente óptimo.

El módem se diseña para las siguientes velocidades:

- velocidad de canal: 4 800 bit/s – velocidad de información 4 800 bit/s,
- velocidad de canal: 4 800 bit/s – velocidad de información 2 400 bit/s (duplicación de la información en pares de subcanales con separación de frecuencia máxima),
- velocidad de canal: 2 400 bit/s – velocidad de información 2 400 bit/s (MDP-1).

Además, a la velocidad de 4 800 bit/s la conmutación a la velocidad de información de 2 400 bit/s se efectúa mediante el código Halley.

El módem puede utilizarse también para la recepción desde dos antenas con diversidad de espacio.

4 Pruebas

4.1 Método

El funcionamiento del circuito radiotelefónico digital se estudió durante un periodo de varios meses en diferentes épocas del año en trayectos de latitudes y meridianos comprendidos entre 1 500-3 000 km y hasta 10 000 km con repetidores).

Se utilizaron dos métodos para la evaluación de la calidad:

- medición de la inteligibilidad de la voz transmitida utilizando tablas de índices de nitidez y de la fiabilidad de la transmisión del tren binario midiendo la proporción de bits erróneos en los intervalos de transmisión de las tablas;
- evaluación subjetiva estadística de la calidad vocal efectuada por los abonados después de una conversación.

En total se realizaron más de 50 mediciones de inteligibilidad y fiabilidad.

4.2 Análisis de los resultados

Se han efectuado determinadas pruebas en distintos periodos del año y diferentes horas del día, durante varios meses en enlaces en ondas decamétricas de una longitud de entre 1 500 y 10 000 km. Durante las pruebas se ha medido, valiéndose de datos regulados, la inteligibilidad silábica del habla transmitida, y se ha evaluado la fiabilidad de transmisión del flujo digital. Se ha efectuado también una evaluación estadística subjetiva de la calidad general del habla, mediante consulta a abonados.

Pueden extraerse las siguientes conclusiones del análisis de los resultados de las pruebas:

- La concentración de errores en un enlace por ondas decamétricas produce una amplia dispersión de la inteligibilidad silábica para una proporción de errores media idéntica.
- La concentración de errores produce un efecto mayor sobre la inteligibilidad de las sílabas con una proporción de error baja (por debajo de 10^{-3}) que con una distribución de errores uniforme. Esto se debe al hecho de que, en el primer caso, los elementos de la tabla de índice de nitidez son afectados, por regla general, no sólo por uno sino por varios errores, con el resultado de que se separan completamente desde el punto de vista del reconocimiento del sonido. Sin embargo, la evaluación del abonado de la calidad global de la conversación con concentración de errores es claramente más favorable que con una distribución de errores uniforme, con la misma densidad media de errores, pues en una conversación establecida un grupo de errores es percibido por el oído como un solo error aislado, con el resultado de que la impresión general es mejor.
- Con una proporción de errores más alta (5×10^{-3} y más), los errores agrupados afectan las secciones individuales de la conversación, pero dejan intactas secciones bastante largas, lo que significa que puede mantenerse un circuito radiotelefónico estable, incluso con una proporción media de errores de aproximadamente 10^{-1} .

- La utilización de doble recepción con antenas con diversidad de espacio aumenta la inteligibilidad silábica recibida en un promedio de 3 a 5% de sílabas y mejora considerablemente la calidad global de la conversación.
- La voz transmitida por radiotelefonía digital es más inteligible que la voz transmitida por un circuito analógico convencional por ondas decamétricas. La calidad global de la voz por radiotelefonía digital es considerablemente mejor cuando se mide con pruebas subjetivas, debido a la ausencia de los efectos característicos del canal en ondas decamétricas, tales como el desvanecimiento selectivo y la interferencia de otras estaciones.
- En el funcionamiento de radioenlaces digitales largos con repetidores, la regeneración de los trenes de bits en el punto de repetición permite mantener un enlace radiotelefónico estable en condiciones en las cuales los métodos normales, incluido Lincompex, no aseguran una calidad satisfactoria de la señal vocal transmitida.
- Las pruebas realizadas en circuitos radiotelefónicos digitales, con vocodificadores diseñados para el funcionamiento a 2 400 y 4 800 bit/s han demostrado que, aunque el vocodificador a 4 800 bit/s produce una calidad de la voz mejor en un canal sin ruido, que el vocodificador a 2 400 bit/s en condiciones reales de circuitos radiotelefónicos por ondas decamétricas, éste último, cuando se combina con un módem que funciona a una velocidad de canal de 4 800 bit/s con reducción a 2 400 bit/s, proporciona en la mayoría de los casos una mejor calidad de la voz. En estas condiciones de funcionamiento, la telefonía digital tiene propiedades de umbral claramente expresadas y la calidad de la señal vocal transmitida con condiciones de propagación deterioradas sigue siendo alta casi hasta el momento en que las comunicaciones se cortan completamente.
- En la radiotelefonía digital, el factor de transmisión de la señal vocal es estable en los componentes de frecuencia generales e individuales. Como los largos retardos de propagación de la señal vocal hacen mucho más audibles las señales de eco, es esencial adaptar un supresor de eco.

ANEXO 4

Sistema de radiotelefonía digital que utiliza un vocodificador de predicción lineal LPC 10

1 Descripción del sistema

El sistema utiliza dos vocodificadores con velocidades de 2 400 bit/s y 800 bit/s.

El primero de ellos, a 2 400 bit/s, proporciona una señal vocal de calidad suficiente como para ofrecer una buena inteligibilidad a velocidad reducida.

La proporción de errores admisible en línea puede llegar a 1 ó 2%, valor aceptable cuando el enlace es de buena calidad.

Si se degrada la calidad del enlace, es decir si la proporción de errores a 2 400 bit/s sobrepasa el 1,5%, se utiliza el segundo vocodificador, a 800 bit/s. Entra entonces en funcionamiento un procedimiento de detección y corrección de errores que sitúa la proporción de errores efectiva a 2 400 bit/s en un valor compatible con la utilización del vocodificador a 800 bit/s, aproximadamente un 1%.

Este segundo vocodificador proporciona una calidad ligeramente inferior a la del vocodificador a 2 400 bit/s, lo que puede considerarse como el compromiso indispensable para poder continuar la comunicación en un canal que prohibiría un enlace con el vocodificador de salida.

Ambos vocodificadores incorporan las siguientes funciones:

- análisis de la señal vocal, para extraer de ésta el conjunto de parámetros que permiten representarla de manera adecuada,
- cuantificación, que transforma los valores digitales de dichos parámetros en flujo digital transmisible,
- posible adición de datos complementarios (redundancia), para detectar y corregir los errores en transmisión,
- descuantificación para regenerar los parámetros originales,
- síntesis, que a partir de los parámetros recibidos reconstituye una señal vocal destinada a producir una impresión acústica, la más próxima posible a la que produciría la señal original, sin intentar reproducir la propia señal.

Las funciones de análisis y síntesis son idénticas para ambos vocodificadores.

Sólo son distintas las funciones de cuantificación, incorporación de redundancia y descuantificación que, cuando actúan con una velocidad en línea de 2 400 bit/s, se refieren a una velocidad útil de 2 400 bit/s u 800 bit/s.

2 Principio de los vocodificadores LPC 10

2.1 Modelado de la señal vocal

En primer lugar, la señal vocal puede considerarse como (casi) estable en el caso de duraciones breves; en esas circunstancias se la puede segmentar en *tramas* de duración constante (en este caso 22,5 ms) en las que se supone que no varían sus características. Por consiguiente, se transmiten a intervalos regulares todos los datos necesarios para sintetizar una o varias tramas de la señal vocal, independientemente de las que preceden o siguen.

A continuación, como la velocidad disponible es muy baja, conviene *modelar* la señal vocal, es decir, representarla por un conjunto de parámetros que reflejen lo más posible la realidad física, o los fenómenos responsables de su producción.

Para ello, se consideran dos casos según que la señal sea periódica (vocal) o no periódica (no vocal).

Para los sonidos periódicos, que corresponden a las vocales, se admite que la señal de palabra se obtiene conformando el espectro de una excitación periódica (mediante filtrados), lo que de hecho representa con bastante exactitud la realidad física en la que las vibraciones acústicas de las cuerdas vocales se propagan en el conducto vocal (laringe, boca, etc.) que desempeña el papel de filtro cuya función de transferencia depende de la vocal pronunciada.

En cuanto a los sonidos no periódicos, existen de dos tipos:

- los sonidos estables o semiestables (del tipo «silbante», como la letra «S»);
- los sonidos transitorios (del tipo «explosivo», como la letra «P»).

Estos dos casos sólo se diferencian por su duración, así como por la velocidad de evolución del nivel sonoro.

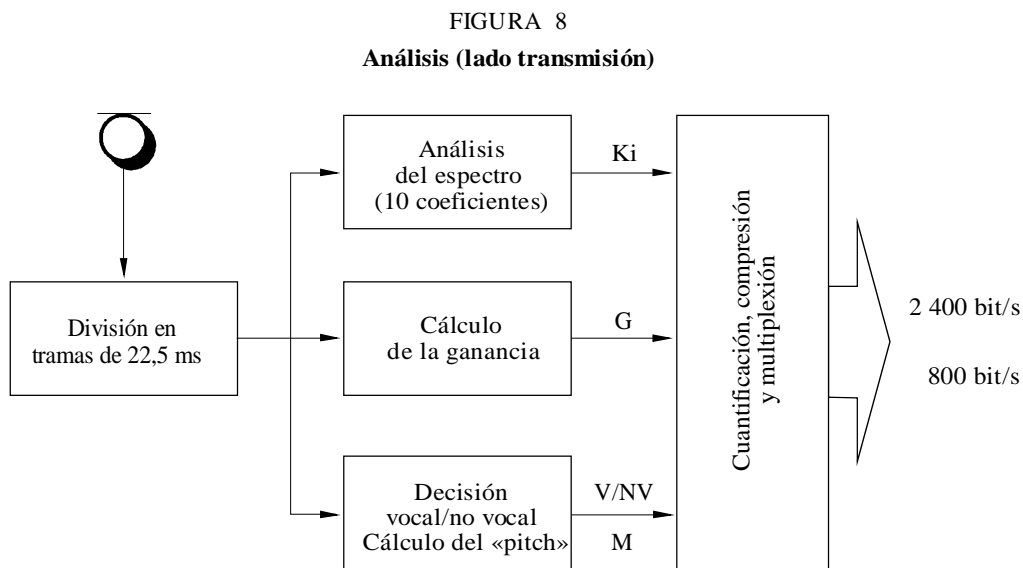
Se les representa a ambos como la salida de un filtro a cuya entrada se aplica una excitación aleatoria (ruido blanco) de evolución más o menos rápida, lo cual se adapta bastante bien a la realidad donde este tipo de sonido no estructurado se produce por turbulencias u oclusiones bruscas en el conducto vocal.

Por último, la señal de palabra en su conjunto siempre se representa como una excitación de naturaleza (periódica o aperiódica) y nivel variables, filtrada de forma que reproduzca lo mejor posible el espectro en frecuencia de la señal original.

El cálculo del filtro se basa en los principios del análisis predictivo, de ahí el nombre de LPC (Linear Predictive Coding – Codificación por predicción lineal) con que se conoce el procedimiento.

2.2 Análisis

Los procesos de análisis extraen de la señal vocal los parámetros necesarios para su modelización y se representan en la Fig. 8.



- Ki: Coeficientes de reflexión
- G: Ganancia para la señal sintética
- V: Sonidos vocalizados (vocales)
- NV: Sonidos no vocalizados (consonantes)
- M: Valor del «pitch» (periodo de reproducción de la señal)

El primer parámetro que debe determinarse es la naturaleza (vocal o no vocal) de la señal. Su evaluación se efectúa basándose en criterios de periodicidad, que varían según el caso. Se calcula en general la autocorrelación a largo plazo de la señal vocal (se observa si a intervalos regulares es idéntica a sí misma), lo que proporciona una primera estimación. Esta estimación se perfecciona siguiendo otros criterios, por ejemplo el nivel instantáneo, la autocorrelación a corto plazo y la relación entre la energía de la señal a frecuencias bajas y a frecuencias altas.

El resultado es un *indicador de vocalización* que toma el valor 1 si se considera que la señal es vocal y 0 en cualquier otro caso.

Cuando se estima que la señal es vocal, conviene calcular su periodo, para traducir su altura. Este periodo, denominado tono o altura («pitch»), se evalúa también en general a partir de la autocorrelación a largo plazo: el intervalo de tiempo tras el cual la señal de palabra se reproduce idéntica a sí misma es el valor de la altura.

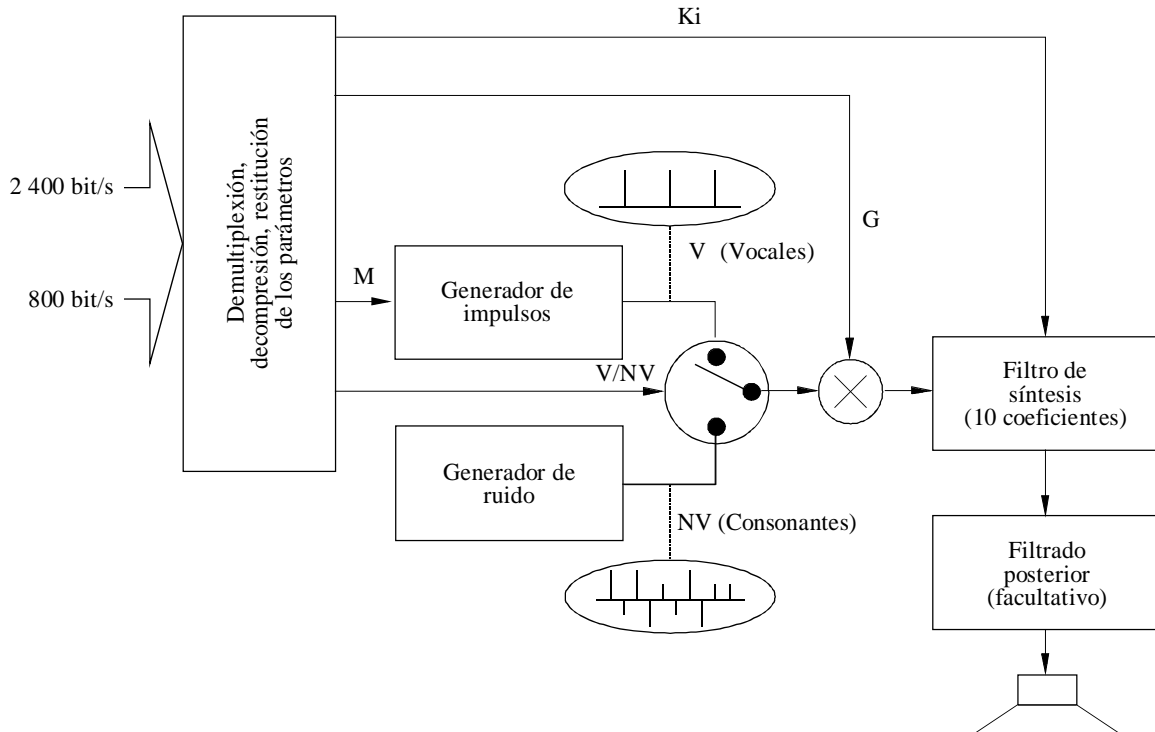
En cuanto al filtro cuya respuesta en frecuencia es la más parecida posible al espectro de la señal de palabra, se calcula a partir de los coeficientes de autocorrelación a corto plazo de los que se deducen los coeficientes del filtro por procedimientos clásicos de tratamiento de señal. El resultado de este análisis es un conjunto de 10 coeficientes de reflexión (de ahí el nombre de LPC 10) que traducen con suficiente fidelidad las variaciones de sección del conducto vocal que, en el origen, ha filtrado (coloreado) la excitación espectral neutra de salida.

Por último, para cada trama, se evalúa el nivel de la señal de palabra a fin de controlar la ganancia del sintetizador situado en el lado síntesis.

2.3 Procedimiento de síntesis

Los algoritmos utilizados para sintetizar la señal vocal reflejan el modelo de producción supuesto de la palabra y se representan en la Fig. 9.

FIGURA 9
Síntesis (lado recepción)



- Ki: Coeficientes de reflexión
- M: Ganancia para la señal sintética
- V: Sonidos vocalizados (vocales)
- NV: Sonidos no vocalizados (consonantes)
- G: Valor del «pitch» (periodo de reproducción de la señal)

En esta figura puede verse:

- un generador de ruido, utilizado para los sonidos no vocalizados;
- un generador de señal periódica, cuya entrada es el «pitch», para los sonidos vocalizados;
- un conmutador, que permite seleccionar uno u otro generador, según la naturaleza de la señal vocal que debe generarse en la trama en curso;
- un filtro de orden 10, que filtra la excitación elegida para proporcionar la coloración; a ese nivel se hace la distinción entre las diferentes vocales y consonantes;
- un sistema de regulación de la ganancia que da a la señal sintética el volumen adecuado;
- opcionalmente, un sistema de «filtrado-posterior» cuyo objeto es enmascarar algunas imperfecciones del sintetizador y mejorar el timbre de la señal sintetizada.

3 Vocodificador a 2 400 bit/s

Para el vocodificador a 2 400 bit/s, los distintos parámetros se cuantifican de manera independiente, trama a trama.

El tono o altura y la vocalización se cuantifican de forma conjunta con 7 bits, el nivel con 5 bits y los 10 coeficientes del filtro con 41 bits para las tramas vocales y algo menos para las tramas no vocales (los bits restantes sirven para supervisar la calidad del enlace).

Tras la incorporación de un bit de tara (sincronización) se obtienen tramas de 54 bits.

Este vocodificador ha sido objeto de un acuerdo de normalización OTAN (STANAG 4198) y de una norma federal en Estados Unidos (Fed. Std. 1015).

4 Vocodificador a 800 bit/s

4.1 Principios utilizados para la reducción de la velocidad

La reducción de la velocidad por un factor de tres se obtiene observando que no sólo los valores de los distintos parámetros están correlacionados entre una trama y la siguiente, sino también que la fidelidad de cuantificación necesaria varía en función del contexto.

Por ejemplo, no es frecuente que el nivel observado en tramas consecutivas varíe mucho; por consiguiente, conviene codificar el nivel por bloques.

Además, si el sonido es estable, también lo es el filtro de predicción y, en consecuencia, es inútil incorporar un nuevo filtro de predicción en cada trama; aunque la velocidad disponible se reduzca por un factor de tres, sigue siendo posible cuantificar con precisión un filtro común para tres tramas consecutivas (por ejemplo). Por el contrario, en una zona transitoria, la propia noción de espectro en frecuencias tiende a difuminarse y consiguientemente puede cuantificarse sin excesiva precisión un filtro por trama, sabiendo que el error de representación cometido vendrá enmascarado por la variación brusca del nivel sonoro.

En consecuencia, el principio básico utilizado para disminuir la velocidad consiste en reagrupar las tramas de la señal vocal en paquetes de tres (por ejemplo) y codificar en bloque cada paquete de parámetros.

4.2 Procedimientos de cuantificación

Cada trama de datos lleva 54 bits representativos de tres tramas de señal vocal. A continuación se describe uno de los procedimientos de cuantificación previstos, que por el momento cuenta con derechos de propiedad.

Se utiliza un total de 10 bits para cuantificar a la vez el «pitch» y la vocalización. Se tiene en cuenta el hecho de que no puede haber más de una transición vocalizada → no vocalizada (o a la inversa) en un paquete de tres tramas y que si hay más de una trama vocalizada, los «pitch» de las tramas adyacentes son vecinos; basta transmitir un valor de referencia y un incremento.

Para el nivel, se utilizan 9 bits, a saber:

- 4 bits para definir un nivel de referencia, común a las tres tramas,
- 5 bits para describir (por lectura de un cuadro denominado diccionario) la evolución del nivel en las tres tramas.

Por último, en lo que respecta al filtro de predicción, se utilizan 35 bits divididos en dos campos.

El primer campo, de 32 bits, está relacionado con:

- un filtro único común a las tres tramas, o
- un filtro común a dos tramas sucesivas y un incremento para obtener el filtro restante, o
- dos filtros.

El segundo campo, de 3 bits, es el encargado de describir el esquema de codificación elegido, entre 8 posibilidades.

El mejor esquema de codificación se elige en el lado de análisis según un criterio de minimización de la distancia espectral ponderada por el nivel de la señal en cada una de las tramas; una trama de nivel bajo no se percibe tan bien como una trama vecina de nivel más elevado.

4.3 Codificación para corrección de errores

El código corrector utilizado para pasar de 800 bit/s a 2400 bit/s es un código de bloques que incorpora 54 bits de información útil por cada paquete de $3 \times 54 = 162$ bits transmitidos.

Puede escogerse, por ejemplo, un código Reed-Solomon reducido a símbolos de 6 bits, que comporta 27 símbolos en total, de los cuales 9 son útiles. Este código, denominado RS(27,9), permite corregir 9 errores en total, es decir una proporción *máxima* del 33% de símbolos con error o una proporción *media* del orden del 20%.

5 Resistencia a los errores de transmisión

La resistencia a los errores de transmisión se efectúa comparando las inteligibilidades respectivas de los dos vocodificadores en función de la proporción de errores a 2400 bit/s, que es la velocidad de transmisión del módem por el canal en ondas decamétricas. Esta inteligibilidad se representa en la Fig. 10.

Puede observarse que es algo más baja a 800 bit/s sin error de transmisión.

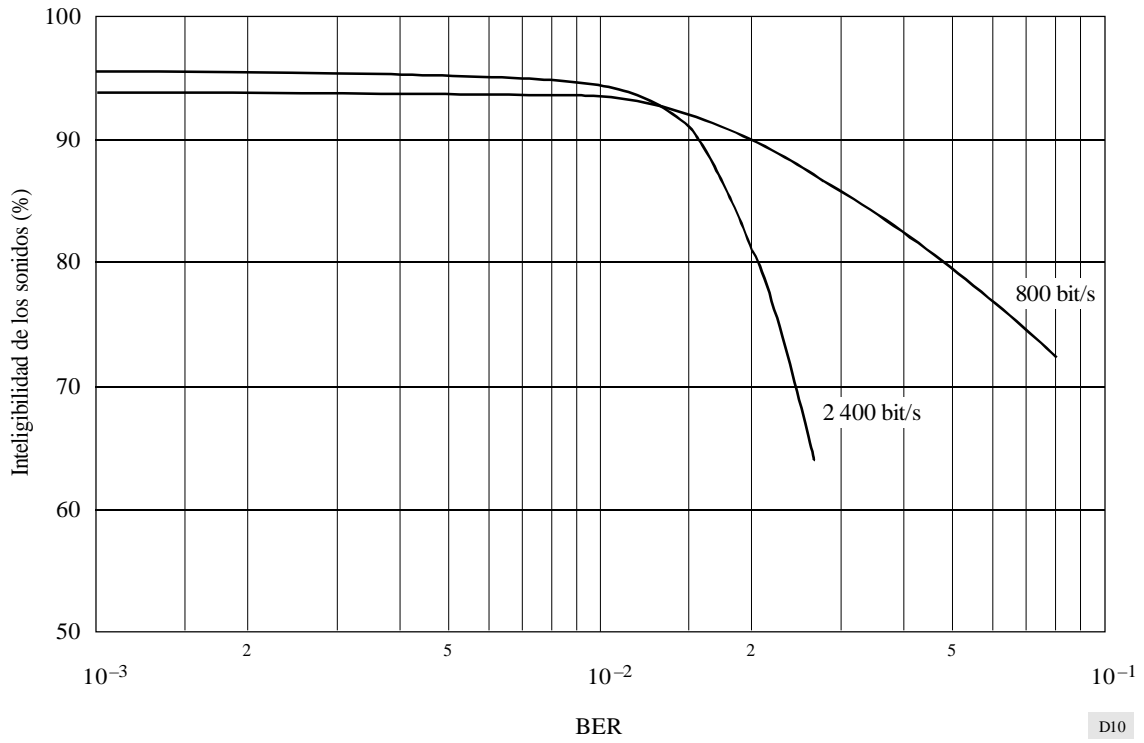
No obstante, cuando la proporción de errores aumenta, la degradación es más lenta que a 2400 bit/s por las dos razones siguientes:

- el código corrector disminuye drásticamente la proporción de errores al pasar de 2400 bit/s a 800 bit/s;
- cuando la proporción de errores es elevada a 2400 bit/s, ello se traduce por tramas suprimidas a 800 bit/s. Estas tramas suprimidas tienen un efecto menor que las tramas falsas.

6 Módem

La transmisión se efectúa por los dos vocodificadores a una velocidad de 2400 bit/s, utilizando un módem normalizado, por ejemplo el módem descrito en la Recomendación UIT-R F.763 o el módem descrito en la norma STANAG 4197.

FIGURA 10
Inteligibilidad de los vocodificadores LPC 10



D10