

UIT-T

SECTOR DE NORMALIZACIÓN
DE LAS TELECOMUNICACIONES
DE LA UIT

G.729

Apéndice II
(08/2005)

SERIE G: SISTEMAS Y MEDIOS DE TRANSMISIÓN,
SISTEMAS Y REDES DIGITALES

Equipos terminales digitales – Codificación de señales
analógicas mediante métodos diferentes de la MIC

Codificación de la voz a 8 kbit/s mediante
predicción lineal con excitación por código
algebraico de estructura conjugada

**Apéndice II: Optimización del anexo B a la
Recomendación G.729 para aplicaciones
de voz por IP – Opción 1**

Recomendación UIT-T G.729 – Apéndice II

RECOMENDACIONES UIT-T DE LA SERIE G
SISTEMAS Y MEDIOS DE TRANSMISIÓN, SISTEMAS Y REDES DIGITALES

CONEXIONES Y CIRCUITOS TELEFÓNICOS INTERNACIONALES	G.100–G.199
CARACTERÍSTICAS GENERALES COMUNES A TODOS LOS SISTEMAS ANALÓGICOS DE PORTADORAS	G.200–G.299
CARACTERÍSTICAS INDIVIDUALES DE LOS SISTEMAS TELEFÓNICOS INTERNACIONALES DE PORTADORAS EN LÍNEAS METÁLICAS	G.300–G.399
CARACTERÍSTICAS GENERALES DE LOS SISTEMAS TELEFÓNICOS INTERNACIONALES EN RADIOENLACES O POR SATELITE E INTERCONEXIÓN CON LOS SISTEMAS EN LÍNEAS METÁLICAS	G.400–G.449
COORDINACIÓN DE LA RADIOTELEFONÍA Y LA TELEFONÍA EN LÍNEA	G.450–G.499
CARACTERÍSTICAS DE LOS MEDIOS DE TRANSMISIÓN	G.600–G.699
EQUIPOS TERMINALES DIGITALES	G.700–G.799
Generalidades	G.700–G.709
Codificación de señales analógicas mediante modulación por impulsos codificados (MIC)	G.710–G.719
Codificación de señales analógicas mediante métodos diferentes de la MIC	G.720–G.729
Características principales de los equipos múltiplex primarios	G.730–G.739
Características principales de los equipos múltiplex de segundo orden	G.740–G.749
Características principales de los equipos múltiplex de orden superior	G.750–G.759
Características principales de los transcodificadores y de los equipos de multiplicación de circuitos digitales	G.760–G.769
Características de operación, administración y mantenimiento de los equipos de transmisión	G.770–G.779
Características principales de los equipos múltiplex de la jerarquía digital síncrona	G.780–G.789
Otros equipos terminales	G.790–G.799
REDES DIGITALES	G.800–G.899
SECCIONES DIGITALES Y SISTEMAS DIGITALES DE LÍNEA	G.900–G.999
CALIDAD DE SERVICIO Y DE TRANSMISIÓN – ASPECTOS GENÉRICOS Y ASPECTOS RELACIONADOS AL USUARIO	G.1000–G.1999
CARACTERÍSTICAS DE LOS MEDIOS DE TRANSMISIÓN	G.6000–G.6999
DATOS SOBRE CAPA DE TRANSPORTE – ASPECTOS GENÉRICOS	G.7000–G.7999
ASPECTOS RELATIVOS AL PROTOCOLO ETHERNET SOBRE LA CAPA DE TRANSPORTE	G.8000–G.8999
REDES DE ACCESO	G.9000–G.9999

Para más información, véase la Lista de Recomendaciones del UIT-T.

Recomendación UIT-T G.729

Codificación de la voz a 8 kbit/s mediante predicción lineal con excitación por código algebraico de estructura conjugada

Apéndice II

Optimización del anexo B a la Recomendación G.729 para aplicaciones de voz por IP – Opción 1

Resumen

El presente apéndice versa sobre la optimización de los algoritmos de detección de actividad vocal (VAD, *voice activity detection*), descritos en el anexo B, y en el cuerpo principal de la Rec. UIT-T G.729 y sus anexos A, C o H, cuando se emplean en aplicaciones de VoIP.

En particular, se tratan los siguientes problemas:

- 1) Funcionamiento inadecuado para señales recibidas con niveles de relación señal/ruido (SNR, *signal-to-noise ratio*) inferiores a 15 dB.
- 2) Ruido molesto similar al de la respiración en la fase CNG.
- 3) Funcionamiento de la VAD en presencia de ruido intenso.
- 4) Inicialización errónea de variables.

Este apéndice está concebido para su utilización en sistemas de VoIP en los que la calidad general se considera más importante que la optimización de la anchura de banda.

Se adjuntan dos ficheros electrónicos. Uno contiene el código fuente modificado en lenguaje C para vad.c (del anexo B/G.729), y el segundo un conjunto de ficheros de prueba que se utilizaron para comprobar el algoritmo que se describe en este apéndice.

Orígenes

El apéndice II a la Recomendación UIT-T G.729 fue aprobado el 5 de agosto de 2005 por la Comisión de Estudio 16 (2005-2008) del UIT-T.

PREFACIO

La UIT (Unión Internacional de Telecomunicaciones) es el organismo especializado de las Naciones Unidas en el campo de las telecomunicaciones. El UIT-T (Sector de Normalización de las Telecomunicaciones de la UIT) es un órgano permanente de la UIT. Este órgano estudia los aspectos técnicos, de explotación y tarifarios y publica Recomendaciones sobre los mismos, con miras a la normalización de las telecomunicaciones en el plano mundial.

La Asamblea Mundial de Normalización de las Telecomunicaciones (AMNT), que se celebra cada cuatro años, establece los temas que han de estudiar las Comisiones de Estudio del UIT-T, que a su vez producen Recomendaciones sobre dichos temas.

La aprobación de Recomendaciones por los Miembros del UIT-T es el objeto del procedimiento establecido en la Resolución 1 de la AMNT.

En ciertos sectores de la tecnología de la información que corresponden a la esfera de competencia del UIT-T, se preparan las normas necesarias en colaboración con la ISO y la CEI.

NOTA

En esta Recomendación, la expresión "Administración" se utiliza para designar, en forma abreviada, tanto una administración de telecomunicaciones como una empresa de explotación reconocida de telecomunicaciones.

La observancia de esta Recomendación es voluntaria. Ahora bien, la Recomendación puede contener ciertas disposiciones obligatorias (para asegurar, por ejemplo, la aplicabilidad o la interoperabilidad), por lo que la observancia se consigue con el cumplimiento exacto y puntual de todas las disposiciones obligatorias. La obligatoriedad de un elemento preceptivo o requisito se expresa mediante las frases "tener que, haber de, hay que + infinitivo" o el verbo principal en tiempo futuro simple de mandato, en modo afirmativo o negativo. El hecho de que se utilice esta formulación no entraña que la observancia se imponga a ninguna de las partes.

PROPIEDAD INTELECTUAL

La UIT señala a la atención la posibilidad de que la utilización o aplicación de la presente Recomendación suponga el empleo de un derecho de propiedad intelectual reivindicado. La UIT no adopta ninguna posición en cuanto a la demostración, validez o aplicabilidad de los derechos de propiedad intelectual reivindicados, ya sea por los miembros de la UIT o por terceros ajenos al proceso de elaboración de Recomendaciones.

En la fecha de aprobación de la presente Recomendación, la UIT ha recibido notificación de propiedad intelectual, protegida por patente, que puede ser necesaria para aplicar esta Recomendación. Sin embargo, debe señalarse a los usuarios que puede que esta información no se encuentre totalmente actualizada al respecto, por lo que se les insta encarecidamente a consultar la base de datos sobre patentes de la TSB.

© UIT 2006

Reservados todos los derechos. Ninguna parte de esta publicación puede reproducirse por ningún procedimiento sin previa autorización escrita por parte de la UIT.

ÍNDICE

	Página
Apéndice II – Optimización del anexo B a la Recomendación G.729 para aplicaciones de voz por IP – Opción 1	1
II.1 Alcance	1
II.2 Referencias	1
II.3 Abreviaturas	2
II.4 Introducción.....	2
II.5 Problemas detectados al utilizar el G.729B en aplicaciones VoIP.....	2
II.6 Diseño experimental.....	3
II.7 Ficheros electrónicos adjuntos	7

Recomendación UIT-T G.729

Codificación de la voz a 8 kbit/s mediante predicción lineal con excitación por código algebraico de estructura conjugada

Apéndice II

Optimización del anexo B a la Recomendación G.729 para aplicaciones de voz por IP – Opción 1*

II.1 Alcance

Si bien en el anexo B a la Rec. UIT-T G.729 se define un mecanismo VAD/DTX/CNG, algunas aplicaciones requieren un algoritmo de detección de la actividad vocal (VAD, *voice activity detection*) diferente, debido a sus limitaciones específicas. Éste es especialmente el caso de las aplicaciones VoIP, en las que el algoritmo descrito en el anexo B muestra un comportamiento inadecuado en las condiciones siguientes:

- 1) Funcionamiento inadecuado para señales recibidas que comienzan con niveles de señal/ruido (SNR) inferiores a 15 dB.
- 2) Ruido molesto similar al de la respiración en la fase CNG.
- 3) Funcionamiento de la VAD en presencia de ruido intenso.
- 4) Por otra parte, la inicialización errónea de variables se examina en el actual anexo B.

El presente apéndice versa sobre las propuestas para corregir los problemas indicados en los puntos 1) a 4) anteriores y constituye una alternativa al actual anexo B.

II.2 Referencias

- [1] Recomendación UIT-T G.729 (1996), *Codificación de la voz a 8 kbit/s mediante predicción lineal con excitación por código algebraico de estructura conjugada*.
- [2] Anexo A a la Recomendación UIT-T G.729 (1996), *Codificador de la voz mediante predicción lineal con excitación por código algebraico de estructura conjugada a 8 kbit/s de complejidad reducida*.
- [3] Anexo B a la Recomendación UIT-T G.729 (1996), *Esquema de compresión de silencios para la Recomendación G.729, optimizado para terminales conformes a la Recomendación V.70*.
- [4] Anexo C a la Recomendación UIT-T G.729 (1998), *Implementación en coma flotante de referencia para la codificación de la voz a 8 kbit/s con CS-ACELP de la Recomendación G.729*.

* Este apéndice consta de dos ficheros electrónicos. Uno contiene el código fuente modificado en lenguaje C para vad.c (del anexo B/G.729), y el segundo un conjunto de ficheros de prueba que se utilizaron para comprobar el algoritmo que se describe en este apéndice.

II.3 Abreviaturas

En este apéndice se utilizan las abreviaturas siguientes.

DSVD	Voz y datos simultáneos digitales (<i>digital simultaneous voice and data</i>)
DTX	Modo de transmisión discontinua (<i>discontinuous transmit mode</i>)
G.729B	Esquema de compresión de silencio (<i>silence compression scheme</i>) definido en el anexo B a la Rec. UIT-T G.729
SID	Descriptor de inserción de silencios (<i>silence insertion descriptor</i>)
SNR	Relación señal/ruido (<i>signal-to-noise</i>)
VAD	Detección de actividad vocal (<i>voice activity detection</i>)

II.4 Introducción

En el anexo B a la Rec. UIT-T G.729 (en adelante, "G.729B") se especifica un esquema de compresión de silencio G.729, optimizado para aplicaciones de voz y datos simultáneos digitales (DSVD) en terminales V.70. Si bien su aplicación inicial era ésta, el G.729B se ha utilizado mucho en aplicaciones VoIP y en la industria seguirán empleándose en el futuro. El G.729B permite a los sistemas de la Rec. UIT-T G.729 (y sus anexos) funcionar en dos modos de transmisión, voz y silencio, que se clasifican mediante la detección de la actividad vocal (VAD). El modo de transmisión discontinua (DTX) sirve para determinar qué tramas de silencio están representados con el descriptor de inserción de silencios (SID). En los últimos años se han comunicado algunos problemas relativos a la utilización del G.729B en aplicaciones VoIP.

II.5 Problemas detectados al utilizar el G.729B en aplicaciones VoIP

En el presente apéndice se trata de resolver los problemas comunicados que se indican a continuación:

- 1) Funcionamiento inadecuado para señales recibidas que comienzan con niveles de señal/ruido (SNR) inferiores a 15 dB.
- 2) Ruido molesto similar a la respiración en las tramas de silencio.
- 3) Funcionamiento del VAD en presencia de ruido intenso.

Problema 1: Funcionamiento inadecuado para señales recibidas que comienzan con niveles de señal/ruido (SNR) inferiores a 15 dB

Si la energía de las 32 tramas iniciales (320 ms) recibidas es inferior a 15 dB, los sistemas G.729B detectan como voz todas las tramas consecutivas que se reciban con energía superior a 15 dB. Esto va en contra del objetivo de utilizar el G.729B, ya que entraña un mayor número de MIPS y mayores costos en memoria y no permite ahorrar anchura de banda.

La solución propuesta consiste en reiniciar el contador de tramas cada 320 ms hasta que las características del ruido de fondo sean inicializadas adecuadamente.

Problema 2: Ruido molesto similar al de la respiración en las tramas de silencio

Cuando después de 129 tramas (1,29 segundos) de tonos u otras señales estacionarias se produce un silencio digital o ruido de muy bajo nivel, la ganancia de ruido en la primera trama SID se calcula a un nivel muy alto, lo que introduce un nivel alto de ruido similar a la respiración y degrada la calidad vocal. Varios clientes de Alcatel y Texas Instruments han informado de este problema. La solución que se propone es introducir un periodo de mantenimiento al final de las tramas vocales.

Problema 3: Funcionamiento del VAD en entornos ruidosos

Cuando la SNR es inferior a 15 dB, el VAD presenta dos problemas: por una parte, la decisión entre voz y ruido oscila, lo que menoscaba el funcionamiento en cuanto al ahorro de anchura de banda debido a actualizaciones del SID y, por la otra, a menudo se toma por ruido una parte de la señal vocal.

En el presente apéndice, la modificación que se introduce al codificador existente es mínima y no se modifica en absoluto el decodificador. El objetivo es perder el mínimo de información posible de la señal vocal, en lugar de optimizar lo más posible la "decisión de ruido".

La solución que se propone reduce la posibilidad de error al decidir entre voz y ruido, especialmente cuando la SNR es pequeña (≤ 15 dB). Consiste en modificar la condición de actualización de las medias móviles de las características de ruido de fondo y la condición de alisamiento de la decisión de actividad vocal inicial.

Las tres modificaciones que se han introducido al codificador existente son las siguientes:

- se elimina la prueba en la que se toma como decisión RUIDO cuando la diferencia entre la señal actual y el ruido anterior es demasiado pequeña;
- se introduce histéresis en la decisión entre VOZ y RUIDO, cuando la energía media del ruido de fondo es lo bastante grande. Se propone que se detecten como mínimo seis decisiones de RUIDO consecutivas antes de pasar de VOZ a RUIDO;
- se modifica la condición necesaria para actualizar las medias móviles de las características del ruido de fondo.

El código C modificado que resuelve estos tres problemas se adjunta en un fichero electrónico al final de este apéndice.

II.6 Diseño experimental

Se han hecho experimentos para verificar la introducción de estas modificaciones.

Problema 1: Funcionamiento inadecuado para señales recibidas que comienzan con niveles de señal/ruido (SNR) inferiores a 15 dB

Las medias móviles de las características del ruido de fondo del VAD del anexo B/G.729 son las siguientes: energía promediada en la totalidad de la banda \bar{E}_f , energía promediada en la banda baja \bar{E}_l , tasa de cruces por cero normalizada \bar{ZC} y parámetros espectrales promediados $\{\overline{LSF}_i\}_{i=1}^p$. Al principio de la función vad_init se ponen a 0 las medias \bar{E}_f , \bar{E}_l , \bar{ZC} y $\{\overline{LSF}_i\}_{i=1}^p$. En el código fuente del anexo B/G.729, el VAD actualiza inicialmente las medias móviles de las características del ruido de fondo utilizando para ello sólo las tramas cuya energía E_f es mayor que 15 dB durante las primeras 32 tramas de la señal de entrada. Cuando la energía de cada una de las primeras 32 tramas es inferior a 15 dB, las medias móviles de las características del ruido de fondo no se actualizan. En la práctica, durante los 320 ms iniciales se lleva a cabo el establecimiento del canal, de modo que el valor del ruido de fondo no corresponde a la realidad. Si el teléfono está en modo silencio o colgado cuando comienza de la conexión del canal, el nivel de señal de entrada es normalmente inferior a 15 dB. En tal caso, las medias móviles de las características del ruido de fondo están a cero y no se efectúa ninguna actualización inicial, lo que se traducirá en que el VAD del anexo B/G.729 detectará como voz todas las tramas de ese canal.

Para mejorar la robustez de la actualización inicial, se propone que el contador de tramas se ponga a cero si la energía de cada una de las 32 tramas es inferior a 15 dB al comenzar el establecimiento de llamada. La versión que se propone del código fuente del anexo B/G.729 es conforme con las disposiciones de los anexos A & B de la Recomendación G.729 actual y con los vectores de prueba del anexo B/G.729.

Problema 2: Ruido molesto similar al de la respiración en las tramas de silencio

Se han hecho pruebas con diferentes tonos a varios niveles (especialmente para los tonos de llamada que se utilizan en Francia y Estados Unidos).

Problema 3: Funcionamiento del VAD en entornos ruidosos

Las pruebas se realizaron en diversas condiciones: con voz clara a niveles nominales, altos y bajos, y con voz acompañada de diversos tipos de ruido de fondo (ruido blanco, de murmullo y de ventilador) a diversos valores de la relación señal/ruido (SNR, *signal-to-noise ratio*).

II.6.1 Las cuatro etapas del alisamiento

El alisamiento de la decisión de actividad vocal se divide en cuatro etapas (véase el § B.3.6 del anexo B/G.729). La cuarta etapa es la siguiente:

si $((E_f < \bar{E}_f + 614)$ y $(Frm_count > 128)$ y $(v_flag = 0)$ y $(rc < 19661))$ entonces (marker = noise)

En primer lugar, esta prueba de alisamiento es la única que no tiene en cuenta las decisiones anteriores, ya que esta función se utiliza para "*reflejar la naturaleza estacionaria a largo plazo de la señal de voz*". Por otra parte, en esta prueba se supone que sea cual fuere el ruido de fondo, siempre habrá al menos 614 entre E y \bar{E}_f , si la trama del caso es una trama de voz. En realidad, cuando la energía del ruido de fondo es demasiado alta, esta afirmación no es verdadera y esta prueba introduce errores en el VAD.

Por consiguiente, la cuarta etapa se ha suprimido. La variable booleana v_flag , que sólo se utiliza en esta prueba, también se ha suprimido.

II.6.2 Histéresis

En un entorno ruidoso, resulta más difícil para la función de decisión de actividad vocal inicial con múltiples límites distinguir claramente entre voz y ruido, dado que los coeficientes del vector de parámetros de diferencia son más pequeños. Por esa razón, el resultado inicial de la VAD contiene más errores en este caso. Para evitar este comportamiento y obtener una señal de salida más coherente, se ha añadido una histéresis al principio de la función de alisamiento:

si (marker = VOICE) entonces (Count_inert = 0)
si ((marker = NOISE) y (Count_inert < 6) y (MeanSE > 8000)) entonces {
 Count_inert ++;
 marker = VOICE; }

En la línea 264 de la tercera etapa se ha añadido la ecuación $Count_inert = 6$ a fin de evitar las interacciones entre las dos pruebas.

II.6.3 La prueba de actualización

De conformidad con la B.3.7/G.729, las medias móviles de las características del ruido de fondo se actualizan si $(E_f < \bar{E}_f + T_6)$. En el código fuente en lenguaje C esta condición se sustituye por la siguiente: $((E_f < \bar{E}_f + 614)$ y $(rc < 24576)$ y $(\Delta S < 83))$.

Considérese el vector de prueba `tstseq3.bin` del anexo B/G.729.

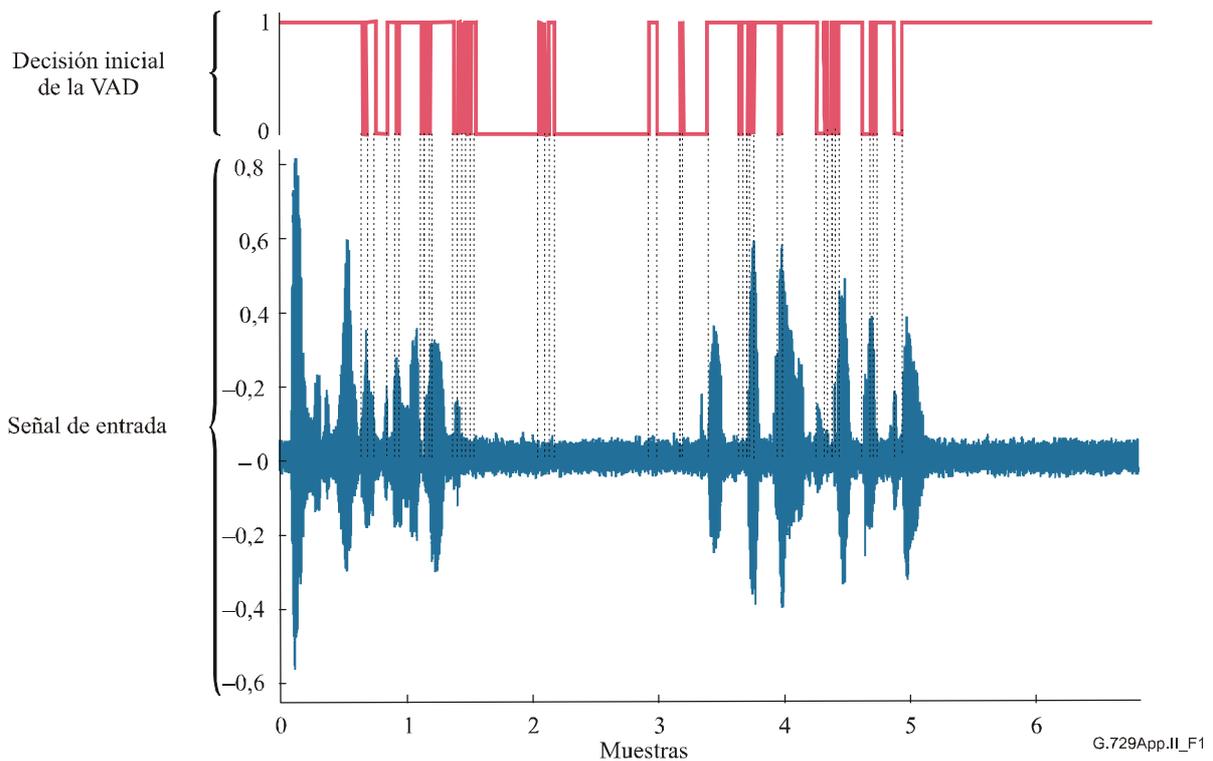


Figura II.1/G.729 – Simulación G.729 + VAD (Código UIT, versión 1.3)

La figura II.1 ilustra la señal de entrada y la decisión inicial adoptada por la VAD con múltiples límites. Se supone que la última parte de la señal, en la que sólo hay ruido, no se transmite. Sin embargo, la decisión inicial de la VAD es constante e igual a uno, aun cuando si el ruido dura mucho tiempo. La causa de este comportamiento es que la condición de actualización descrita antes sigue siendo errónea, dado que el valor de ΔS es demasiado alto. En la figura II.2 se representa ΔS para la prueba actual (línea continua de la curva roja) y sin la condición ($SD < 83$) (línea de puntos de la curva azul).

NOTA – La distorsión espectral ΔS se representa mediante SD en el código fuente en lenguaje C.

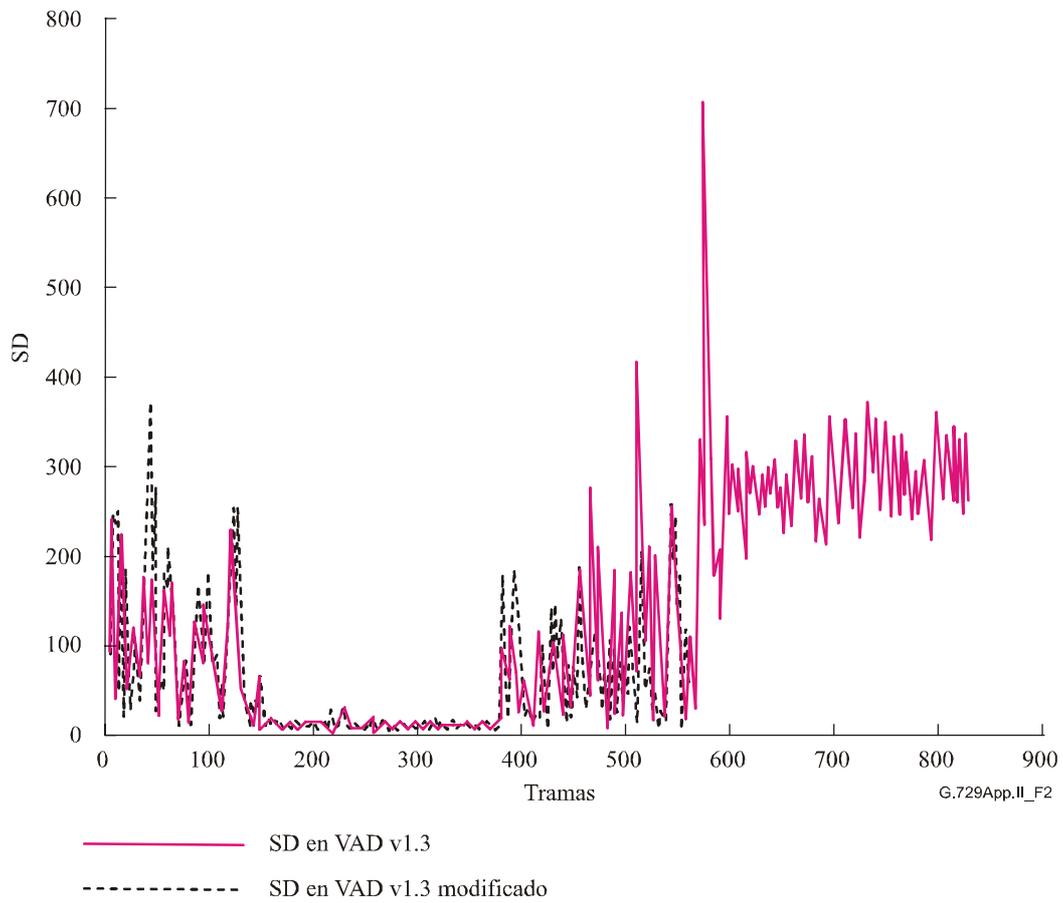
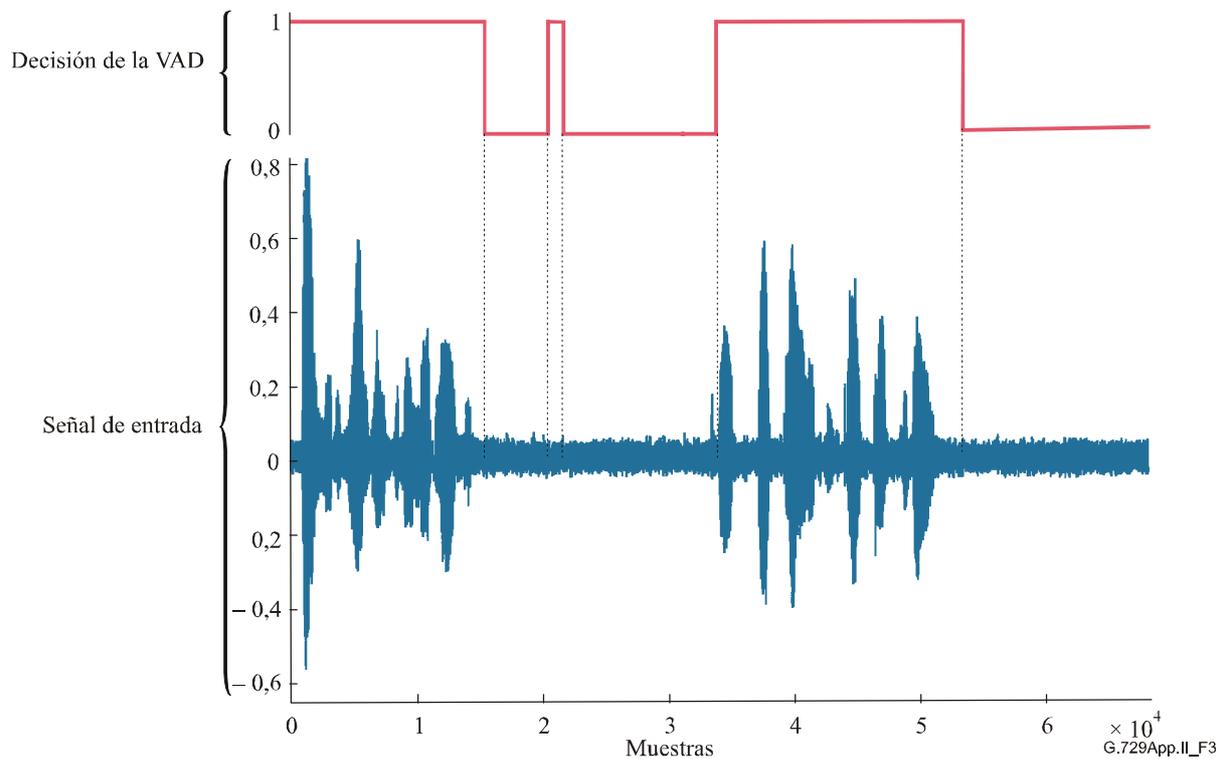


Figura II.2/G.729 – Comparación de valores SD

Sin esta condición, las medias móviles se actualizan correctamente y el valor de ΔS es mucho menor, como supuestamente es el caso en un entorno ruidoso. La decisión adoptada por el algoritmo VAD también es correcta (véase la figura II.3).



**Figura II.3/G.729 – G.729 + VAD
(Código UIT, versión 1.3, modificada)**

II.7 Ficheros electrónicos adjuntos

Se adjuntan a este apéndice dos conjuntos de ficheros electrónicos.

El primero contiene el fichero `vad.c` actualizado, que implementa las modificaciones del algoritmo de la VAD descrita en el anexo B/G.729, de acuerdo con lo indicado en este apéndice.

El segundo fichero adjunto contiene un conjunto de ficheros de prueba empleados para comprobar el algoritmo descrito en este apéndice.

SERIES DE RECOMENDACIONES DEL UIT-T

Serie A	Organización del trabajo del UIT-T
Serie D	Principios generales de tarificación
Serie E	Explotación general de la red, servicio telefónico, explotación del servicio y factores humanos
Serie F	Servicios de telecomunicación no telefónicos
Serie G	Sistemas y medios de transmisión, sistemas y redes digitales
Serie H	Sistemas audiovisuales y multimedios
Serie I	Red digital de servicios integrados
Serie J	Redes de cable y transmisión de programas radiofónicos y televisivos, y de otras señales multimedios
Serie K	Protección contra las interferencias
Serie L	Construcción, instalación y protección de los cables y otros elementos de planta exterior
Serie M	Gestión de las telecomunicaciones, incluida la RGT y el mantenimiento de redes
Serie N	Mantenimiento: circuitos internacionales para transmisiones radiofónicas y de televisión
Serie O	Especificaciones de los aparatos de medida
Serie P	Calidad de transmisión telefónica, instalaciones telefónicas y redes locales
Serie Q	Conmutación y señalización
Serie R	Transmisión telegráfica
Serie S	Equipos terminales para servicios de telegrafía
Serie T	Terminales para servicios de telemática
Serie U	Conmutación telegráfica
Serie V	Comunicación de datos por la red telefónica
Serie X	Redes de datos, comunicaciones de sistemas abiertos y seguridad
Serie Y	Infraestructura mundial de la información, aspectos del protocolo Internet y Redes de la próxima generación
Serie Z	Lenguajes y aspectos generales de soporte lógico para sistemas de telecomunicación