



## UNIDAD DIDÁCTICA 3



### **FORMATOS DE AUDIO Y VIDEO. Calidad de Aplicación**

20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

1



## Objetivos



- Conocer y entender los formatos de transmisión para streaming de audio y vídeo.
- Conocer las formatos contenedores.
- Conocer los aspectos más importantes de calidad de aplicación de servicios streaming.

20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

2



## Competencias adquiridas



- Capacidad de entender y diseñar los formatos de transmisión de streaming que se aplican en la actualidad.
- Comprender la evaluación del funcionamiento de estos sistemas y su interacción con los niveles inferiores.
- Conocer y entender los formatos contenedores de los sistemas audiovisuales.

20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

3



## Contenidos



- **UNIDAD DIDÁCTICA 3: FORMATOS DE AUDIO Y VIDEO**
  - **AUDIO.**
    1. Voz: G.711 ... G.729.
    2. Encapsulado RTP.
  - **VIDEO.**
    1. H.261, H.263 y H.264.
    2. Encapsulado RTP.
  - **CONTENEDORES.**
    - 1.- MP3.
    - 2.- MPEG4.
  - QoA. PSNR y PSQM, MOS.

20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

4



## Bibliografía



- *O. Hersent, D. Gurle IP Telephony: Packet-Based Multimedia Communications Systems Addison-Wesley, Hardcover, December 1999. ISBN 0201619105*  
*Standares ITU: P.800, P.860, P.861.*

- 72G711 ITU-T Recommendation G.711: *Pulse Code Modulation (PCM)* of voice frequencies. 1972.
- 90G726 ITU-T Recommendation G.726: 40, 32, 24, 16 kbps *Adaptive Differential Pulse Code Modulation (ADPCM)*. December 1990.
- 92G728 ITU-T Recommendation G.728: Coding of Speech at 16 kbps using Low-Delay Code Excited Linear Prediction. September 1992.
- 93H261 ITU-T Recommendation H.261: Video Codec for Audiovisual Services at p x 64 kbits. March 1993.
- 95BT601 ITU-R Recommendation BT.601-5: Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios. October 1995.
- 96G723 ITU-T Recommendation G.723.1: Dual Rate Speech Codec for Multimedia Communications Transmitting at 5.3 and 6.3 kbps. March 1996.
- 96G729 ITU-T Recommendation G.729: C source code and test vectors for implementation verification of the G.729 8 kbps CS-ACELP speech coder. March 1996.
- 96H263 ITU-T Recommendation H.263: Video Coding for Low Bit Rate Communication. March 1996.
- 98H263 ITU-T Recommendation H.263: Video Coding for Low Bit Rate Communication. January 1998.
- 98T38 ITU-T Recommendation T.38: Procedures for real-time Group 3 facsimile communication over IP networks. June 1998.
- 99T38 ITU-T Recommendation T.38 (Amendment 1): Procedures for real-time Group 3 facsimile communication over IP networks. April 1999.
- 99T140 ITU-T Recommendation T.140: Protocol for multimedia application text conversation. February 1998 with addendum 1 to T.140 in February 2000.
- 00G711A2 ITU-T Recommendation G.711, Appendix II: A comfort noise payload definition for ITU-T G.711 use in packet-based multimedia communication systems. February 2000.
- G723A96 ITU-T Recommendation G.723.1, Annex A: C reference code, test signals and test sequences for the fixed point 5.3 and 6.3 kbps dual rate speech coder and for the silence compression scheme, version 5.1. November 1996.
- G729A96 ITU-T Recommendation G.729, Annex A: C source code and test vectors for implementation verification of the G.729 reduced complexity 8 kbps CS-ACELP speech coder. November 1996.
- G729B97 ITU-T Recommendation G.729, Annex B: C source code and test vectors for implementation verification of the algorithm of the G.729 silence compression scheme. August 1997.

20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

5



## Audio codec



- Codec de audio implica pérdidas.
- En IP hay pruebas subjetivas y objetivas de calidad de aplicación
  - PSQM - Medición de la Calidad de percepción del habla (UIT-T P.861) representa fielmente el criterio humano y la percepción comparación algorítmica entre la señal de salida y una entrada para saber tipo de altavoz, el volumen, la demora, los marcos de activo / silencio, la saturación, el ruido ambiental.
- Objetivo: buen MOS 4.0.

20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

6



## Audio codec



Condiciones de la voz:

- Modo speech:
  - Aire empujado desde los pulmones pasando las cuerdas vocales y en el tracto vocal
  - Las vibraciones básicas - las cuerdas vocales
  - El sonido se ve alterado por la disposición del tracto vocal (lengua y la boca)
- Modelo del tracto vocal como un filtro
  - El cambio de la forma de onda es relativamente lento
  - La señal de excitación modeliza las vibraciones en las cuerdas vocales

20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

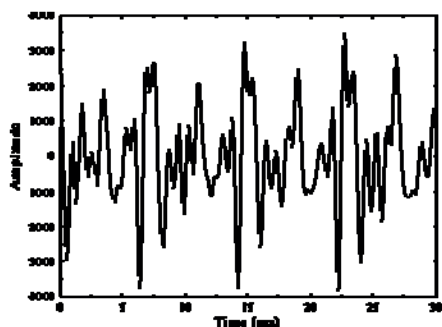
7



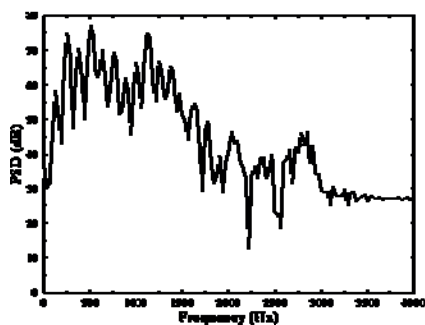
## Audio codec



• Voz (Hablado)



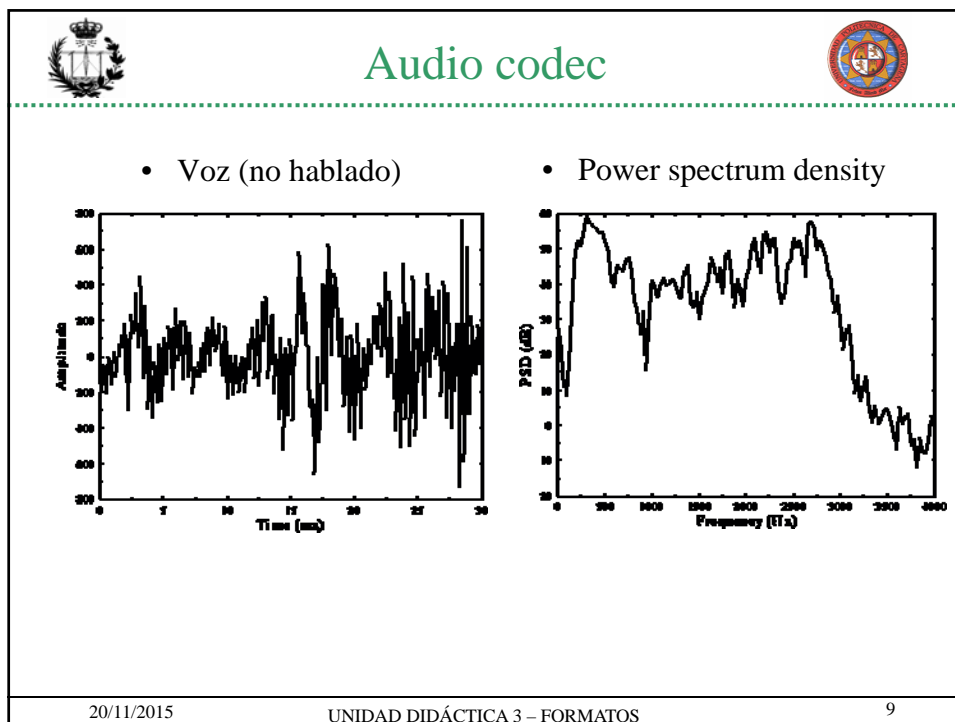
• Power spectrum density




20/11/2015


UNIDAD DIDÁCTICA 3 – FORMATOS

8






## Audio codec




---

- Codecs Wavelenghts
  - Muestrea, cuantifica y codifica
  - De alta calidad y no complejo
  - Gran cantidad de ancho de banda
- Codecs de fuente (vocoders)
  - Modeliza la señal de entrada a un modelo matemático
  - Modelos de filtros lineales de predicción del tracto vocal
  - Una secuencia sonora / sorda para la excitación controlada por flags
  - Se envía una determinada información en lugar de la señal
  - Se generan bajas tasas de bits, pero los sonidos son sintéticos
  - Los valores más altos no mejoran mucho

20/11/2015
UNIDAD DIDÁCTICA 3 – FORMATOS
10

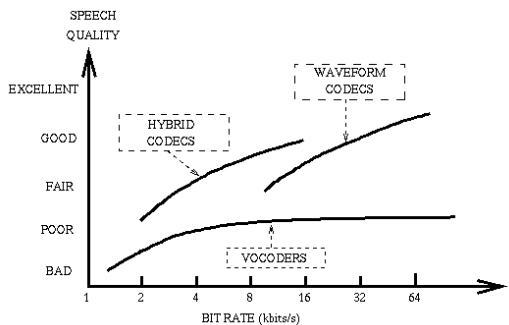


## Audio codec




---


- **Codecs Híbridos (Hybrid)**
  - Intento de ofrecer lo mejor de ambos
  - Realiza un cierto grado de coincidencia de formas de onda
  - Utiliza el modelo de producción de sonido
  - Calidad bastante buena en velocidad de bits baja




20/11/2015
UNIDAD DIDÁCTICA 3 – FORMATOS
11



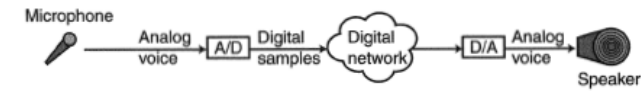
## Voz: G.711



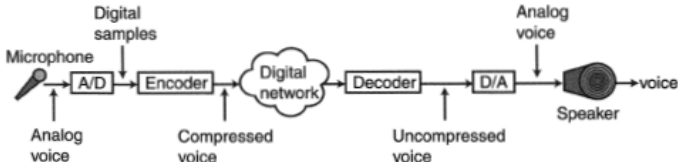
---



Microphone



Microphone



20/11/2015
UNIDAD DIDÁCTICA 3 – FORMATOS
12



## Voz: G.711



**G.711.** The G.711 coder is specified in [72G711] and is widely used on digital telecommunications networks. In North America and Japan, G.711 is known as  $\mu$ -law, and in Europe it is known as A-law. G.711  $\mu$ -law and A-law compress 14-bit linear **Pulse-Code Modulation (PCM)** samples and 13-bit linear PCM samples, respectively, into 8-bit logarithmic samples—producing a bit rate of 64 kbps.

Equación de codificación:

$$F(x) = \text{sgn}(x) \begin{cases} \frac{A|x|}{1+\ln(A)}, & |x| < \frac{1}{A} \\ \frac{1+\ln(A|x|)}{1+\ln(A)}, & \frac{1}{A} \leq |x| \leq 1, \end{cases}$$

En Europa,  $A = 87.7$ ; el valor 87.6 también se usa.

Ecuación de decodificación:

$$F^{-1}(y) = \text{sgn}(y) \begin{cases} \frac{|y|(1+\ln(A))}{A}, & |y| < \frac{1}{1+\ln(A)} \\ \frac{\exp(|y|(1+\ln(A))-1)}{A}, & \frac{1}{1+\ln(A)} \leq |y| < 1. \end{cases}$$

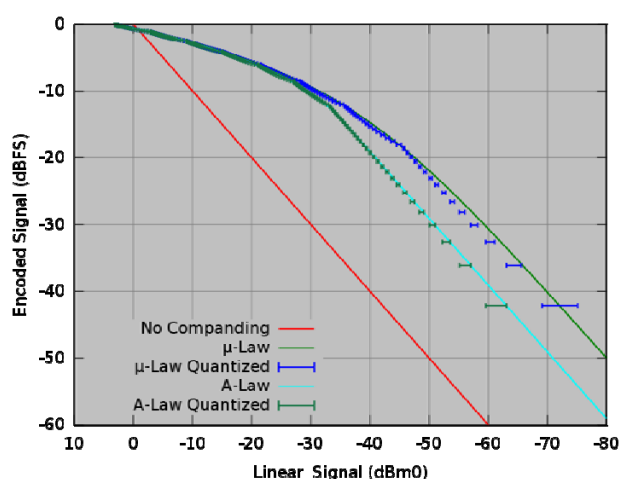
20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

13



## Voz. G.711.



20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

14



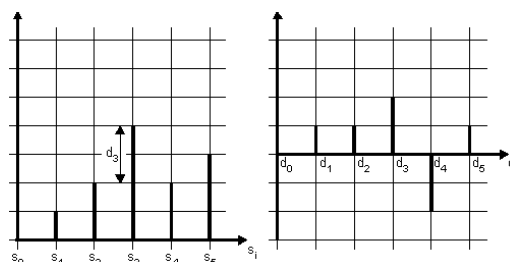
## Voz. G.711.



G.711 variante: Codificación diferencial:

DPCM, Differential PCM:

- Sólo transmiten la diferencia entre el valor predicho y el valor real
- Cambios en la voz de forma relativamente lenta
- Es posible predecir el valor de una muestra de función de los valores de las muestras anteriores
- El receptor debe realizar la misma predicción
- La forma más simple es Ninguna Predicción
- No hay retraso algorítmico



20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

15



## Voz. G.711.



ADPCM, Adaptive DPCM

Predice valores de la muestra sobre la base de muestras anteriores

Basado en algún conocimiento de cómo el “habla” varía con el tiempo

El error está cuantificado y se transmite

Menor número de bits necesarios

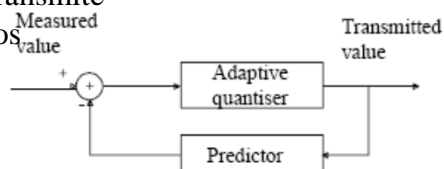
Aplicada en G.721

32 kbps

Aplicada en G.726

A-law/mu-law PCM -> 16, 24, 32, 40 kbps

Se obtiene un MOS alrededor 4.0 a 32 kbps




20/11/2015


UNIDAD DIDÁCTICA 3 – FORMATOS

16





## Voice codecs



---

Hybrid codecs

Llenar la diferencia entre los códecs de forma de onda y la fuente

La más exitosa y de uso común

Dominio del tiempo-ABS codecs

No es un simple modelo de dos estados, voz / silencio

Diferentes señales de excitación son soportadas.


Más cercano a la forma de onda original se selecciona:

**MPE**, Excitación Multi-Pulse


**RPE**, Excitación Regular-Pulse

**CELP**, Excitación con código lineal predictivo

20/11/2015
UNIDAD DIDÁCTICA 3 – FORMATOS
17



## Voice Codecs



---

CELP codecs, características:

- Un filtro cuyas características cambian con el tiempo
- Un **libro** de códigos de vectores acústicos
- Un **vector** es un conjunto de elementos que representan caracteres diferentes de la fuente
- Transmitir los coeficientes del filtro, la ganancia, un valor instantáneo para cada muestra.


Low Delay (LD-CELP )

- Adaptación codificador hacia atrás
- Utiliza las muestras anteriores para determinar los coeficientes de filtro
- Opera con cinco muestras a la vez, sólo la primera se transmite
- Delay <1 ms
- 1024 vectores en el libro de códigos
- 10-bit indicador (índice)
- 16 kbps


Nota sobre el LD-CELP codificador

Minimiza el error cuadrático medio

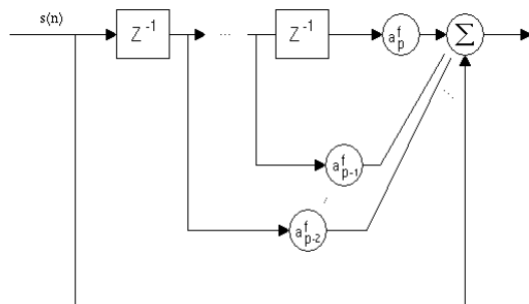
20/11/2015
UNIDAD DIDÁCTICA 3 – FORMATOS
18




## Voice codecs




Synthesis Filtro:

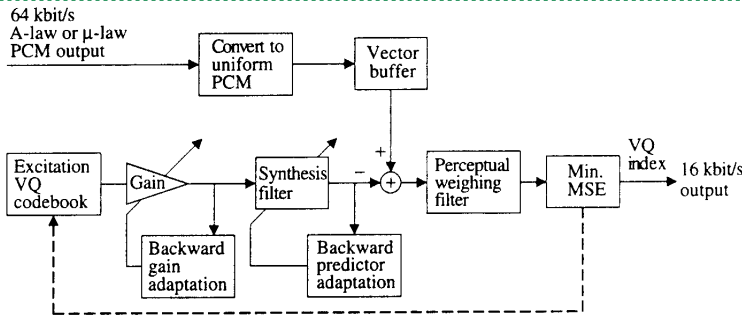


01/12/2011
UNIDAD DIDÁCTICA 3 – FORMATOS
19

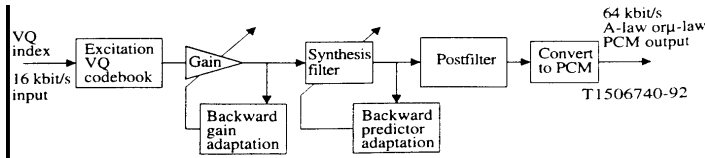


## Voice codecs





**LD-CELP decoder**



*b) LD - CELP decoder*

Un MOS de alrededor de 3.9 para el ancho de banda un cuarto de G.711

20/11/2015
UNIDAD DIDÁCTICA 3 – FORMATOS
20



## Voz. G.723.1.



**G.723.1.** The G.723.1 coder is specified in [96G723] and uses the **Multi-Pulse Maximum Likelihood Quantization (MP-MLQ)** algorithm to compress 30 ms blocks of 240 16-bit samples into 24-byte frames, producing a bit rate of **5.3** kbps. Another algorithm, **Algebraic Code-Excited Linear Prediction (ACELP)**, compresses 30 ms blocks to 20-byte frames—producing a bit rate of **5.3** kbps. G.723.1 produces one of three frame sizes: 24 bytes, 20 bytes, or four bytes. The 4-byte frame specifies comfort noise parameters. There is no restriction on how 4-, 20-, and 24-byte frames are intermixed. The least significant 2 bits of the first byte in the frame determine the frame size and codec type as follows: bits 00 imply a 24-byte frame; bits 01 imply a 20-byte frame; bits 10 imply a 4-byte frame; and bits 11 are reserved for future use. We can switch between the two rates at any 30 ms frame boundary. G.723.1 also specifies the VAD and CNG algorithms in [G723A96].

20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

21



## Voz. G.723.1.



### Características:

- Frecuencia de muestreo 8 kHz/16 bits (240 muestras de 16 bits por cada 30 ms de audio).
- Tasa de bits fija (5,3 kbps tramas de 20 bytes comprimiendo bloques 240 muestras de 16 bits de 30 ms; 6,3 kbps con 24 bytes de bloques 240 muestras de 16 bits de 30 ms).
- Retraso algorítmico es de 37,5 ms por trama.
- La complejidad del algoritmo tiene una potencia de **25**, utilizando una escala relativa, en referencia a G.711; mientras G.729a es 15 [G.729a es menos complejo que G.723].
- G.723.1 Anexo A define 4 bytes silencio bajo la inserción del descriptor (SID) -> comfort noise.
- **PSQM (medida de percepción de calidad de voz) en pruebas de rendimiento de las condiciones ideales producía una media de puntuaciones de opinión (MOS) de 4,08 para la G.723.1 (6,3 kbps), en comparación con un valor original de 4,45 por G.711.**
- **PSQM (medida de percepción de calidad de voz) en pruebas de rendimiento de la red de estrés daba MOS de 3.57 para G.723.1 (5,3 kbps), en comparación con 4.13 para G.711.**

20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

22



## Voz. G.723.1.



El codificador ACELP

- Una banda limitada para la entrada de señal de voz
- Muestreada a 8 kHz, 16 bits de cuantización uniforme PCM
- Operar en los bloques de 240 muestras a la vez
- Obtiene ganancia con el método look-ahead de 7,5 ms
- Un retraso total de algoritmos de 37,5 ms + otros retrasos
- Un filtro de paso alto para eliminar cualquier componente de CC
- Diversas operaciones para determinar los coeficientes de un filtro adecuado
- 5,3 kbps, Código Algebraico con predicción lineal
- 6,3 kbps, múltiples pulsos de cuantificación de máxima verosimilitud
- La transmisión es de coeficientes lineales de la predicción de parámetros de ganancia
- Excitación indexados codebooks
- 24-octetos por tramas a 6,3 kbps, 20 bytes a 5,3 kbps

20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

23



## Voz. G.723.1.



G.723.1 Anexos A

Silence Insertion Description (SID) tramas de 4 octetos

Los dos LSBs del primer octetos indican

00 **6.3kbps** 24 octetos/trama

01 **5.3kbps** 20

10 SID frame 4

**Un MOS de alrededor 3.8**

Al menos retardos 27.5 ms

20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

24



## Voz. G.726.




G.726. The G.726 coder is specified in [90G726]. This coder uses the *Adaptive Differential Pulse Code Modulation (ADPCM)* algorithm to compress 8-bit samples of 64 kbps G.711 to 2-bit, 3-bit, 4-bit, or 5-bit samples at the bit rate of 16, 24, 32, or 40 kbps, respectively. You can convert samples from G.726 back to G.711 samples.




## Voz. G.728.



G.728. The G.728 coder is specified in [92G728]. This coder uses the *Low Delay* —Code-Excited Linear Prediction (LD-CELP) algorithm to compress 0.625 ms blocks of five 16-bit samples into 10-bit samples at the bit rate of 16 kbps.




## Voz. G.729.




---

G.729. The G.729 coder is specified in [96G729] and [G729A96]. This coder uses the *Conjugate Structure—Algebraic Code-Excited Linear Prediction (CS-ACELP)* algorithm to compress 10 ms blocks of 80 16-bit samples into one of two frame sizes: 10-byte (8 kbps) or two-byte. The two-byte frame specifies comfort noise parameters. G.729 also specifies the VAD and CNG algorithms in [G729B97].

20/11/2015
UNIDAD DIDÁCTICA 3 – FORMATOS
27



## Voz. G.729.



---

- Ancho de banda 8 kbps (usual)
- Comprime bloques formados por 80 muestras de 16 bits en 10 bytes (8kbps) o 14 bytes (11.2kbps)
- 5 ms look-ahead
- Retraso algorítmico de 15 ms
- Un códec complejo
- G.729.A (Anexo A), una serie de simplificaciones
  - Estructura de la trama misma
  - Codificador / decodificador, G.729/G.729.A
  - Calidad ligeramente inferior

20/11/2015
UNIDAD DIDÁCTICA 3 – FORMATOS
28



## Voz. G.729a.



Las características de G.729a son:

- Frecuencia de muestreo 8 kHz/16 bits (comprime 80 muestras en bloques de 10 ms)
- Tasa de bits fija (8 kbps)
- Tamaño de trama fijo (10 bytes)
- Retraso algorítmico es de **15 ms** por trama, con **5 ms** retraso de pre-análisis
- G.729a es un codificador híbrido que utiliza el código algebraico emocionado de predicción lineal (ACELP).
- La complejidad del algoritmo tiene un valor de **15**, utilizando una escala relativa en función de G.711; siendo G.723.1 de **25**.
- **Pruebas PSQM en los rendimientos de las condiciones ideales: MOS promedio del 4,04 para G.729a**, en comparación con 4,45 de G.711 (u-law)
- **Pruebas PSQM en los rendimientos de tráfico normal de la red: MOS del 3,51 por G.729a**, en comparación con 4.13 para G.711 (u-law).

20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

29



## Otros codecs




CDMA QCELP se define en el ES-733

- Codificador de tasa variable
- Dos tipos más comunes:
  - La tasa alta de 13,3 kbps
  - Una tasa más baja, 6,2 kbps
- Supresión de silencio
- Para su uso con RTP ver RFC 2658


20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

30



## Otros codecs




---

- GSM Enhanced Full-Rate (EFR)
  - Una versión mejorada de GSM full-rate
  - Utiliza el codec ACELP
  - 12,2 kbps
  - Soporte de transmisión discontinua
  - Para su uso con RTP ver RFC 1890


20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

31



## Otros codecs



---

- GSM Adaptive Multi-Rate (AMR) codec
  - ✓ Retardo de codificación de 20 ms
  - ✓ Ocho modos diferentes: desde 4,75 kbps a 12,2 kbps
  - ✓ 12,2 kbps, GSM EFR
  - ✓ 7,4 kbps, IS-641 (TDMA sistemas celulares)
  - ✓ Se puede cambiar el modo en cualquier momento
  - ✓ Oferta transmisión discontinua
    - ✓ El SID (Descriptor Silencio) se envía tras cada bloque y es de 5 bytes de tamaño
  - ✓ La opción de codificación de muchas redes inalámbricas 3G

20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

32

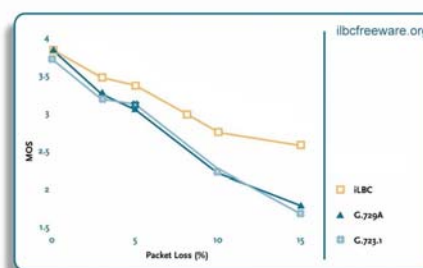




## Otros codecs, iLBC



- Un codificador libre para VoIP robusta
- 13.33 kbit/s con una codificación de tramas cada 30 ms y 15.20 kbps para 20 ms
- Complejidad computacional en el rango de G.729A (medio-bajo)



The tests were performed by Dynstat, Inc., an independent test laboratory.  
Score system range: 1 = bad, 2 = poor, 3 = fair, 4 = good, 5 = excellent

20/11/2015

UNIDAD DIDÁCTICA

Courtesy of GLOBAL IP SOUND



## Comparativa




- Para los valores MOS actuales medidos en laboratorio
  - G.711 no funciona para paquetes perdidos
  - G.729 se puede acomodar a paquetes perdidos por interpolación de paquetes anteriores
    - Pero causa error en las siguientes tramas
- Coste de procesado
  - G.728/ G.729, 40 MIPS
  - G.726 10 MIPS


20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

34



## Atributos




---


**Bit rate.** A lower bit rate is important from at least two standpoints. First, it is economical to squeeze more voice streams into the same bandwidth network. Second, in multimedia applications, the savings in bit rate from voice can be allocated to other streams such as video or data. It is no coincidence that the bit rate of **G.723.1 is the lowest among the codecs** mentioned here. G.723.1 was primarily developed for videophones where the channel bandwidth is shared between voice and video.

**Complexity.** The number of *Millions of Instructions per Second* (MIPS) that is needed to execute the encoder and the decoder algorithms measures the complexity of the codec. For portable applications, the usage of *Read Only Memory* (ROM) and *Random Access Memory* (RAM) is also important, but for PC applications, memory is of little concern. More complexity translates to higher *Central Processing Unit* (CPU) and memory usage, resulting in more cost and power usage. Measuring the complexity of a coder depends on a number of factors, and chief among them are the processor being used, the memory requirements, and the implementation of the algorithm. **The G.728 codec is the most complex—about 15 times more complex than G.726 and about twice as complex as G.723.1. On the other hand, G.711 is the simplest and takes a negligible amount of processing power.**

20/11/2015
UNIDAD DIDÁCTICA 3 – FORMATOS
35



## Atributos




---


**Delay.** The delay that is introduced by the coder depends on three factors: frame size, look-ahead, and processing delay. The coders work on sets of samples called frames. **G.723.1 has a frame size of 30 ms,** which is equal to 240 samples, and **G.729 has a frame size of 10 ms** which is equal to 80 samples. In addition to waiting for a complete frame, some encoders also wait to collect some samples ahead of the frame for analysis in compressing the frame. This additional waiting time is the look-ahead time. The look-ahead time for **G.723.1 is 7.5 ms,** and for **G.729 it is 5 ms.** Delay due to frame size and the look-ahead is called the algorithmic delay. Algorithmic delay for **G.723.1 is 37.5 ms,** and for **G.729 it is 15 ms.** Other encoders in Table 2.1 do not have look-ahead. G.711 and G.726 are sample-based coders with a frame size of one sample.

The algorithmic delay is constant for the given codec, but the processing delay depends on the processing power of the hardware that you use. The processing delay also depends on the complexity of the codec. The delay is equal to the time that it takes to encode and decode one frame. Thus, codec delay equals the algorithmic delay plus the processing delay.

20/11/2015
UNIDAD DIDÁCTICA 3 – FORMATOS
36




## Atributos.




---

**Quality.** The quality of the codec is commonly measured through the *Mean Opinion Score* (MOS). This rating is subjective, where 1 is bad and 5 is excellent. Experts assign an MOS score by comparing the quality of voice from a specific codec with other known codecs. The MOS score of G.711 is 4, which is the quality of voice on standard telephone lines (also known as toll quality). Both G.729 and the high bit rate version of G.723.1 have a near-toll quality.

20/11/2015
UNIDAD DIDÁCTICA 3 – FORMATOS
37




## Attributes.




---

| Codec   | Bit rate in kbps | Complexity compared with 6.726 | Algorithmic Delay | Quality MOS |
|---------|------------------|--------------------------------|-------------------|-------------|
| G.711   | 64               | very low                       | 0.125             | 4.0         |
| G.723.1 | 5.3              | 8                              | 37.5              | 3.9         |
| G.723.1 | 6.3              | 8                              | 37.5              | 3.6         |
| G.726   | 32               | 1                              | 0.125             | 3.85        |
| G.728   | 16               | 15                             | 0.625             | 3.61        |
| G.729   | 8                | 10                             | 15                | 3.9         |
| G.729A  | 8                | 6                              | 15                | 3.7         |

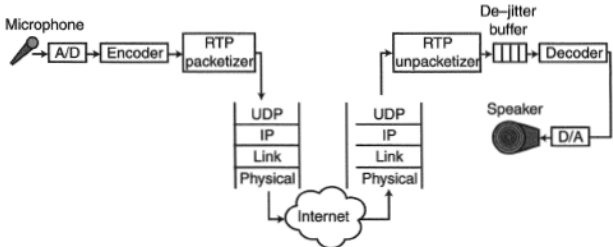
20/11/2015
UNIDAD DIDÁCTICA 3 – FORMATOS
38



## RTP format.




---




UDP MTU (Maximum transmission unit) 512 bytes.  
Sin fragmentación.  
Retardo End-to-End => TCP no es válido => UDP.

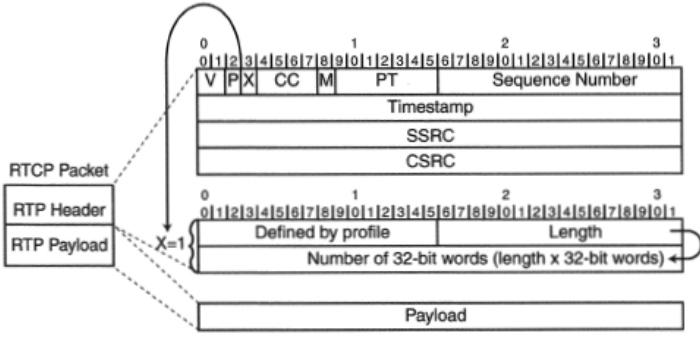
20/11/2015
UNIDAD DIDÁCTICA 3 – FORMATOS
39



## RTP format.



---



20/11/2015
UNIDAD DIDÁCTICA 3 – FORMATOS
40



## RTP format: header



- *Version (two bits)*. This field provides the version of RTP that the application is using. RFC 1889 is version 1 [CAS96].
- *Padding (two bit)*. This field aligns the network packet to a boundary that the link layer or the application requires. If the padding bit is set, the last byte in the RTP packet contains the count of the number of redundant bytes at the end of the RTP packet to be ignored.
- *Extension (one bit)*. This field extends the RTP header. Some implementations use this for experimentation purposes. Other implementations that are not aware of the specifics of the extension can ignore it. If this field is set, the header is extended by two more fields: Defined By Profile and Length. The interpretation of Defined By Profile is specific to the implementation. The Length field provides the number of 32-bit words that follow. The Extension field is normally set to zero (its use is rare).

20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

41



## RTP format: header



- *Marker (one bit)*. The RTP profile in RFC 1890 specifies the use of this bit for voice and video applications. For voice applications, this bit is set only in the first packet following silence suppression. Note that during silence suppression, packets are not generated. The receiver can use this indication to dynamically adjust the delay in the play-out buffer. For video applications, this bit is set to one in the last packet of a video frame; otherwise, it is set to zero.
- *Payload Type (seven bits)*. The RTP profile in RFC 1890 specifies the use of this field to identify the format of the RTP payload. The 128 payload type numbers are divided into static and dynamic payload types as follows: 0–34 is assigned statically; 35–71 is currently unassigned; 72–76 is reserved; 77–95 is unassigned; and 96–127 is assigned dynamically. A static payload type is permanently assigned to the payload format. Examples are a payload type of 4 for G.723.1 and 34 for the 1996 version of H.263. The conference control protocol assigns the dynamic payload type. The trend is to do away with static payload types and have payload types assigned only dynamically. Static payload types are no longer assigned to new payload formats that are developed. The 1998 version of H.263, also known as H.263+, does not have a static payload type.

20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

42



## RTP format: header



- **Sequence Number (16 bits).** This field increases by one for each RTP packet that is sent on the network. The receiver uses this field to detect packet loss and to correctly order any out-of-order packets.
- **Timestamp (32 bits).** Three clocks are used in the end-to-end transmission of media: a sampling clock to sample the analog media, an RTP clock to generate the timestamp, and a reference clock to correlate the timestamps from separate media for the purpose of synchronizing the media streams. The timestamp reflects the instant at which the first sample in the payload was sampled. The timestamp is derived from a clock that increments monotonically and linearly with sufficient resolution to enable various calculations (such as synchronization and packet arrival jitter). The frequency of the RTP clock depends on the payload format. The RTP clock for voice payload formats normally equals the sampling clock. The voice payload formats use the RTP clock from the set 8000Hz, 11025Hz, 16000Hz, 22050Hz, 24000Hz, 32000Hz, 44100Hz, and 48000Hz. Video payload formats use an RTP clock of 90000Hz, which yields exact integer timestamp increments for most widely used frame rates—24Hz for *High-Definition Television* (HDTV), 25Hz (PAL), 29.97Hz (NTSC), and 30Hz (HDTV). All packets belonging to the same video frame have the same timestamp. Video frames are distinguished by a change in timestamp. In the layered scenario, all frames corresponding to the same temporal reference use the same timestamp.

20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

43



## RTP format: payload



The payload G723.1 is composed by:

- The frequency of the RTP clock is the same as the sampling clock of 8000Hz.
- The payload consists of one or more G.723.1 frames. The number of frames in the payload is signaled through the H.323 Control Protocol
- A static payload type of 4 is assigned.

The payload format for comfort noise is specified in [SCHU96] and includes the following components:

- The frequency of the RTP clock is 8000Hz.
- The payload consists of a single byte whose most significant bit is set to zero, and the least-significant seven bits contain the absolute noise level expressed in -dBov. This payload provides an absolute noise level between 0 and -127 dBov.
- A static payload type of 13 is assigned.

20/11/2015

UNIDAD DIDÁCTICA 3 – FORMATOS

44



## Cuestiones de clase



- 1.- Describir las mejoras/propiedades/inconvenientes de los distintos formatos de transmisión de audio.
- 2.- Calcula el retardo mínimo de los distintos formatos de compresión de audio vistos en esta unidad.
- 4.- El bit **marker** de la cabecera RTP marca el tiempo necesario para calcular el End-to-End Delay (para calcular el tiempo de reproducción apropiado) pero cómo resolverías que se perdiera el paquete con el **marker** activado.
- 5.- Ofrece un ejemplo completo de realización de una llamada telefónica.