Servicios Multimedia
Grado de Ingeniería en Tecnología de Telecomunicación – 4º curso

**Resumen de investigación:** Audio (Voz)

Elige dos o tres de los conceptos descritos en clase (voz, codecs de voz, compresión de voz, G.711, archivos de audio, contenedores de audio,..), realiza una investigación, y escribe un párrafo en el que describes los elementos de valor que has identificado.

---

**Concepto 1 (Codecs de Voz) :**

Codecs de Voz: Objetivos y Tipos Especializados

Los codecs de voz son herramientas diseñadas para comprimir y descomprimir señales de audio enfocadas en la voz humana. La importancia de estos codecs radica en que buscan optimizar la calidad de la transmisión de voz en aplicaciones de telecomunicaciones, videollamadas y otros servicios de voz sobre redes. A diferencia de los codecs de audio en general, los codecs de voz tienen un enfoque específico en preservar la claridad y la naturalidad de las señales vocales, lo cual es esencial para mantener la comunicación eficaz. En este contexto, algunos codecs son especializados para distintos tipos de audio, como la música y la voz.

Para el caso de la música, codecs como Vorbis son más adecuados, ya que se enfocan en mantener la fidelidad en un amplio rango de frecuencias. Esto permite que el sonido sea nítido y detallado, características importantes para la reproducción de música. Sin embargo, cuando el objetivo es la voz, existen codecs especializados como GSM y Speex, que optimizan la transmisión de la voz humana a través de canales de baja calidad o baja tasa de bits. Estos codecs están diseñados para reproducir claramente las características del habla, asegurando que el mensaje sea inteligible y que se mantengan elementos esenciales para la identificación del hablante.

Los objetivos principales de los codecs de voz son dos. Primero, la inteligibilidad del mensaje, es decir, que el contenido de la conversación sea claro y entendible para el receptor. Esto es esencial en aplicaciones donde la precisión de la comunicación es prioritaria. Segundo, la identificación del hablante, que permite que el oyente pueda reconocer quién está hablando, ya que la voz humana tiene características únicas que ayudan en este reconocimiento, como el timbre, el tono y ciertas inflexiones particulares de cada individuo.

Técnicas y Comparación de Codecs de Voz

La compresión de voz se basa en técnicas específicas que buscan reducir la redundancia de la señal de audio. Esto incluye la eliminación de correlaciones a corto y largo plazo que ocurren naturalmente en el habla. Por ejemplo, la correlación a corto plazo, que ocurre aproximadamente cada 1 milisegundo, y la correlación a largo plazo, que sucede en intervalos de 5 a 10 milisegundos, se eliminan para reducir la cantidad de datos necesarios para representar la señal.

Existen varias técnicas para la compresión de voz, principalmente clasificadas en waveform, parametric y hybrid. Los codecs waveform intentan replicar directamente la forma de onda del audio original y suelen ser utilizados cuando la fidelidad de la señal es prioritaria. Por otro lado, los

parametric codecs, también conocidos como vocoders, son más eficientes para el habla, ya que utilizan modelos de producción de voz humana para recrear los sonidos, eliminando redundancias inherentes al habla. Finalmente, los hybrid codecs combinan elementos de ambos enfoques, tratando de obtener un equilibrio entre calidad y eficiencia en la compresión, lo cual es particularmente útil en aplicaciones de voz sobre IP y sistemas de telecomunicaciones.

Cada técnica tiene ventajas y desventajas dependiendo del contexto de uso. Los codecs waveform suelen requerir más datos y ancho de banda, pero mantienen una alta fidelidad. Los parametric codecs, en cambio, son más compactos y eficientes en términos de ancho de banda, pero pueden perder naturalidad en la voz.

Bibliografía

*Speex: a free codec for free speech*. (s. f.). https://www.speex.org/

*ar5iv – Articles from arXiv.org as responsive HTML5 web documents*. (s. f.). Ar5iv. https://ar5iv.labs.arxiv.org/

*XIPHWiki*. (s. f.). https://wiki.xiph.org/Main_Page

Valin, J. (s. f.). *Jean-Marc Valin*. https://jmvalin.ca/

---

**Concepto 2 (Compresión de Voz) :**

Técnica Basada en Forma de Onda para Compresión de Voz

La compresión de voz basada en forma de onda es un método directo de codificación de la señal de voz que se centra en reducir la redundancia en la forma de onda sin realizar un análisis profundo de las características del habla. En este método, se eliminan redundancias en la señal de voz y se utiliza la reconstrucción de la onda para lograr compresión. Es una técnica de baja complejidad, lo que significa que su implementación es sencilla y requiere menos recursos de procesamiento. Las tasas de bits obtenidas suelen ser moderadas, entre 16 kbps y 64 kbps, y se utilizan en aplicaciones donde se necesita calidad en tiempo real sin demasiado consumo de ancho de banda. Entre los ejemplos de codecs de este tipo están el PCM (Pulse Code Modulation) y el ADPCM (Adaptive Differential Pulse Code Modulation), este último adaptado tanto para banda estrecha (NB) como para banda ancha (WB).

Técnica Paramétrica en la Compresión de Voz

La compresión paramétrica de voz se basa en un análisis detallado de los segmentos de la señal de voz, típicamente en intervalos de aproximadamente 20 milisegundos. En este enfoque, cada segmento se clasifica como sonoro o no sonoro, y se extraen parámetros específicos, como el tono, la frecuencia y la amplitud de cada segmento. Estos parámetros son luego codificados y transmitidos en lugar de la señal de onda completa, lo cual permite una mayor compresión. Sin embargo, esta técnica requiere una alta complejidad de procesamiento, dado que necesita algoritmos avanzados para analizar y clasificar con precisión los parámetros de cada segmento. Si bien esta técnica ofrece

mejores ratios de compresión que el enfoque de forma de onda, la calidad resultante puede ser inferior, especialmente para frecuencias altas. Un codec representativo de este tipo es el LPC (Linear Prediction Coding), que opera a tasas de bits bajas, entre 1,2 y 4,8 kbps, siendo especialmente útil en aplicaciones de comunicaciones inalámbricas seguras, como comunicaciones militares.

Técnica Híbrida en la Compresión de Voz

La compresión de voz híbrida o técnica de codificación “Analysis-by-Synthesis” (análisis por síntesis) combina aspectos tanto de la forma de onda como de la compresión paramétrica. Este enfoque permite reconstruir la señal de voz con alta fidelidad y, al mismo tiempo, aprovechar los beneficios de la compresión basada en parámetros. En este método, la señal original se analiza y luego se ajusta hasta obtener una representación que conserve la calidad del habla de manera eficiente. La técnica híbrida es más compleja que la de forma de onda y requiere algoritmos iterativos que mejoran la calidad y reducen la tasa de bits. El codec CELP (Codebook Excitation Linear Prediction) es un ejemplo de este método, con una tasa de bits de 4,8 a 16 kbps. CELP es ampliamente utilizado en comunicaciones móviles, redes satelitales y VoIP, donde se busca una calidad de voz cercana a la de la telefonía convencional (con un MOS superior a 4.0). Otros codecs híbridos modernos, como G.729, G.723.1, AMR, iLBC y SILK, también utilizan esta técnica para optimizar la calidad en redes de ancho de banda limitado.

Bibliografía

*Speex: a free codec for free speech*. (s. f.-b). https://www.speex.org/

*SpEEx: a free codec for free speech*. (s. f.). Ar5iv. https://ar5iv.labs.arxiv.org/html/1602.08668

---

**Concepto 3 (G.711) :**

G.711: Codec de Referencia para Compresión de Voz

El codec G.711 es un estándar en la compresión de voz que se conoce como "PCM (Pulse Code Modulation) de frecuencias de voz". Desarrollado por la ITU-T, G.711 está diseñado para transmitir señales de voz en redes de telecomunicaciones, como las redes de telefonía pública conmutada (PSTN) y Voz sobre IP (VoIP). Utiliza una tasa de muestreo de 8 kHz (8000 muestras por segundo) con una profundidad de 8 bits por muestra, lo que da una tasa de bits de 64 kbps. Esta alta tasa de bits permite una calidad de voz muy cercana a la voz en tiempo real, haciéndolo ideal para servicios de telefonía que buscan alta fidelidad en la transmisión de voz.

Cuantificación de la Voz y Codificación Logarítmica No Uniforme

G.711 emplea una técnica de cuantificación logarítmica no uniforme para aprovechar las características de la voz humana. Dado que las señales de voz tienen una función de densidad de probabilidad (PDF) que varía, este codec asigna más niveles de cuantificación a los sonidos de menor amplitud, que ocurren con mayor frecuencia en el habla humana. Este método de cuantificación logarítmica permite que las señales de bajo nivel (como susurros o pausas) se reproduzcan con mayor precisión que las señales de alto nivel, optimizando así la calidad percibida sin aumentar la tasa de bits.

Variaciones del G.711: μ-law y A-law

Existen dos variaciones del codec G.711, adaptadas para diferentes regiones:

- μ-law (Mu-law): Utilizado en Norteamérica y Japón, este esquema de codificación transforma una señal de 14 bits en una de 8 bits. La codificación μ-law es conocida por su precisión en niveles bajos de señal y su compresión en niveles altos, lo que mejora la relación señal-ruido en sistemas de telefonía.
- A-law: Utilizado en Europa, A-law convierte una señal de 13 bits en una señal de 8 bits, proporcionando un rango dinámico ligeramente menor que μ-law, pero ofreciendo una mejor calidad en niveles medios de señal. A-law es particularmente eficaz en ambientes donde se prioriza la consistencia en la calidad de voz, como en redes telefónicas europeas.

Bibliografía

*Speex: a free codec for free speech*. (s. f.-c). https://www.speex.org/

*SpEEx: a free codec for free speech*. (s. f.-b). Ar5iv. https://ar5iv.labs.arxiv.org/html/1602.08668

*Speex FAQ - XiphWiki*. (s. f.). https://wiki.xiph.org/Speex_FAQ