

CAP 3. COMPRESSÃO DE DADOS MULTIMÍDIA

Aula 7: Padrões de compressão multimídia – Codec de Voz

INE5431 Sistemas Multimídia
Prof. Roberto Willrich (INE/UFSC)
roberto.willrich@ufsc.br

Compressão de Dados Multimídia

Conteúdo:

- Necessidade de compressão
- Princípios da compressão
- Classificação das técnicas de compressão
- Medição do desempenho de compressão
- Técnicas de compressão sem perdas
 - RLE, Huffman, LZW (GIF), Codificação Preditiva
- Técnicas de compressão de áudio, vídeo e imagens
- Padrões de compressão multimídia
 - JPEG, Codecs de Voz, MPEG, MPEG-4, H.261, H.263

Codecs e Quadro de Voz

Quadro de voz

- A maioria dos codificadores de voz se baseia em quadros
 - Quadros de voz ou pacotes de voz
 - Cada quadro de voz tem uma duração de 1 a 30 ms
- Codecs compactam quadros de voz
 - Contendo um número fixo de amostras
 - Número de amostras depende do codec utilizado

Codecs e Quadro de Voz



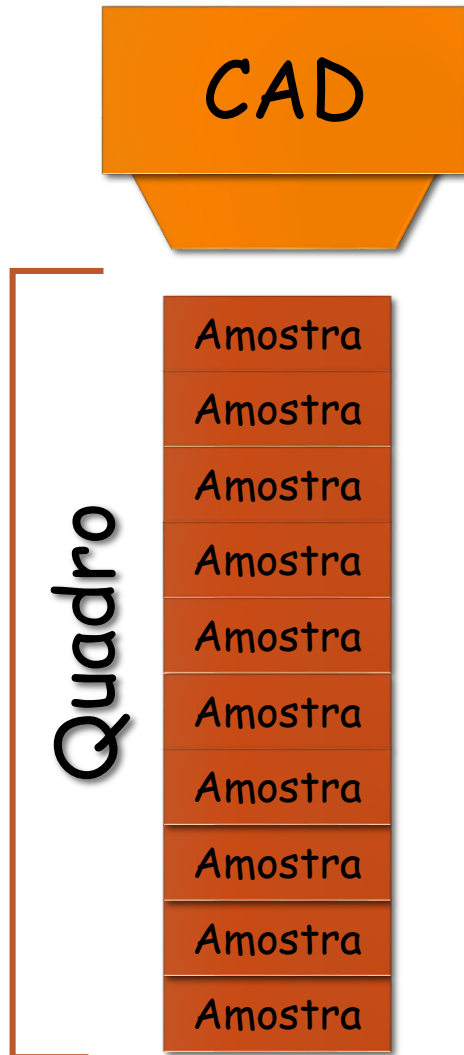
CAD

Quadro

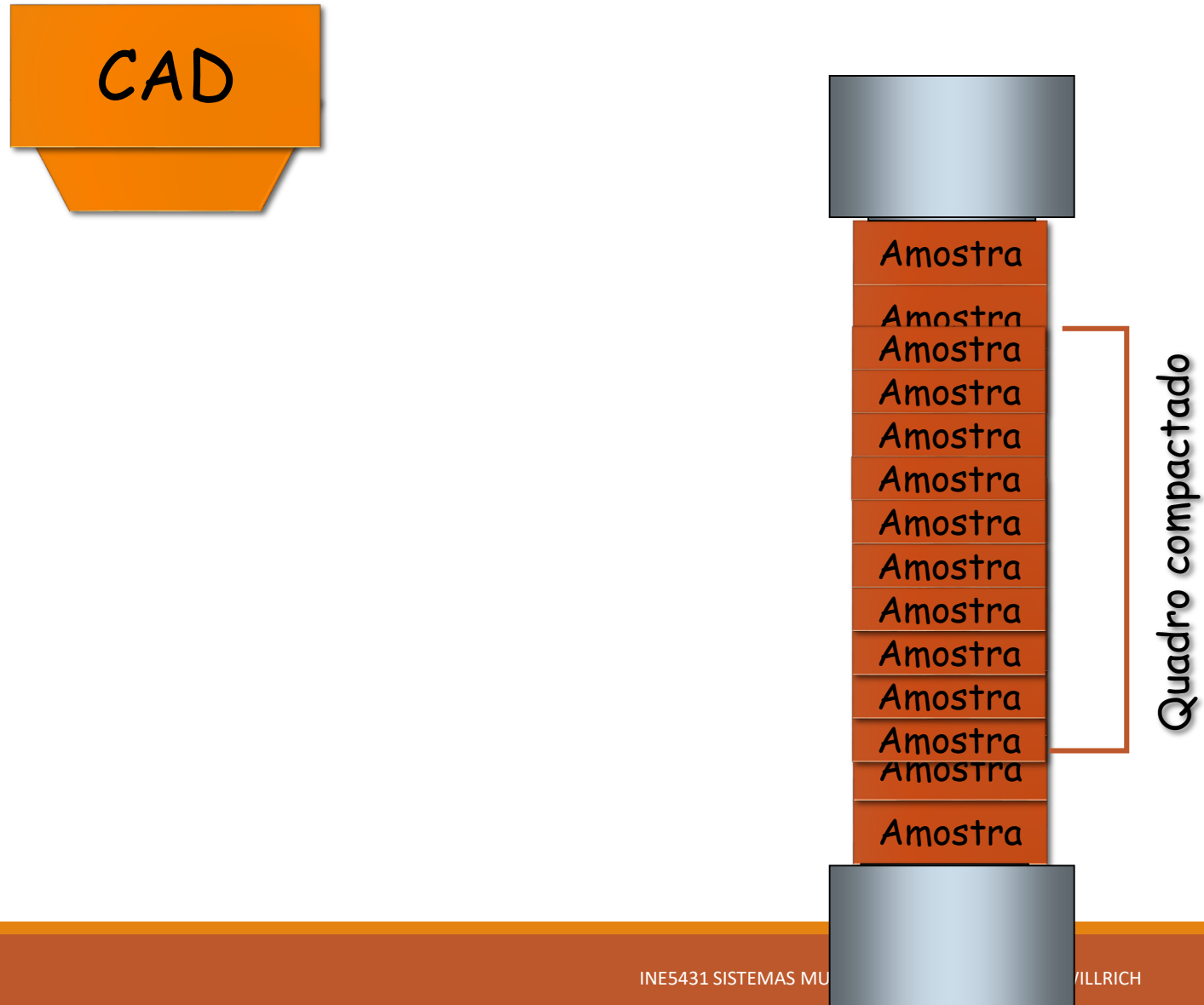
Montagem do quadro

- Fluxo de dados de áudio precisa ser acumulado
 - até que ele atinja o tamanho do bloco antes de ser processado pelo codificador
- Acumulação de amostra leva tempo
 - soma-se ao atraso fim-a-fim.

Codecs e QUADRO de Voz



Codecs e QUADRO de Voz



Codecs de Voz

Codecs recomendados pela ITU-TS para Voz

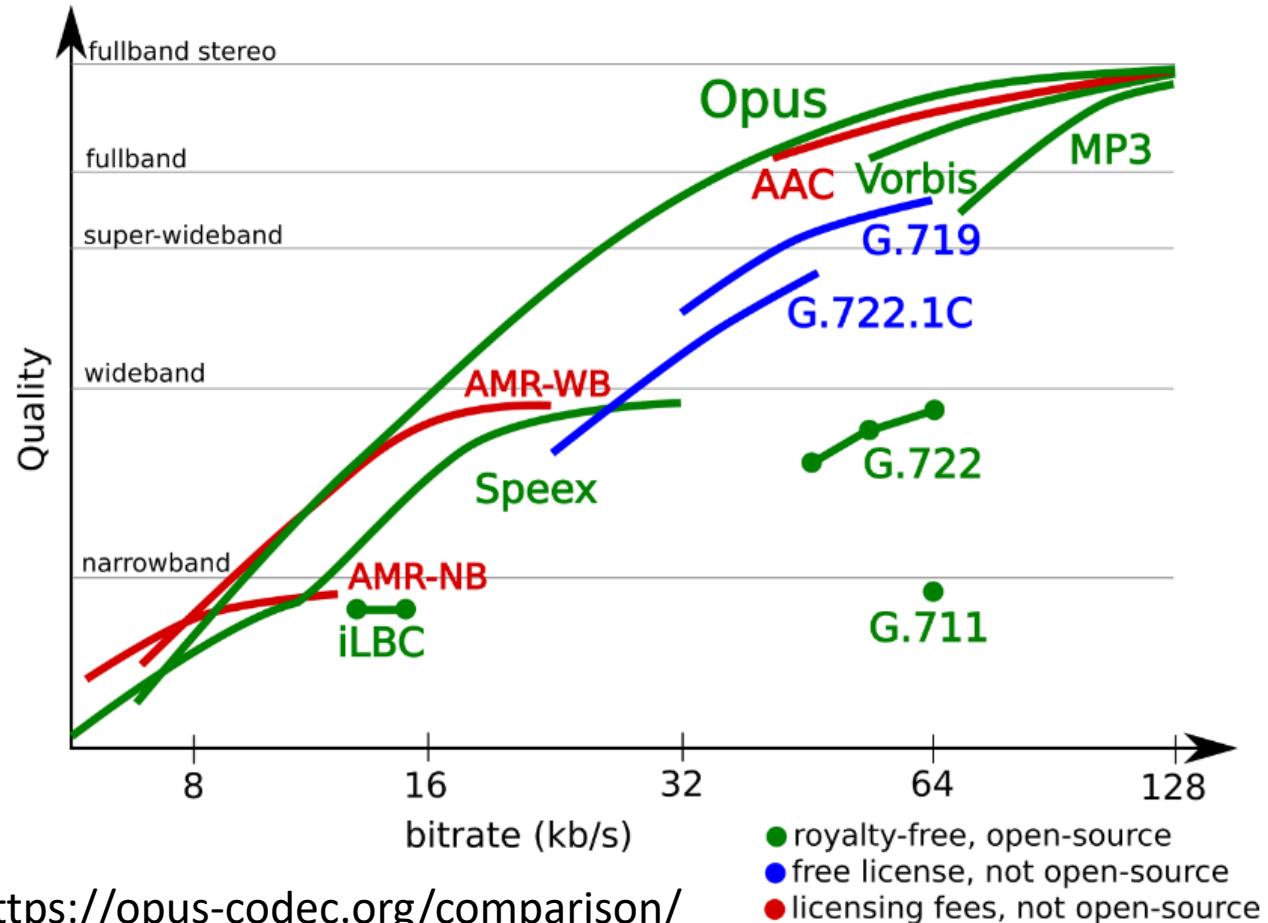
- Codecs clássicos de voz, como G.711, G.721, G.722, G.723,...

Outros codecs

- Opus, sucessor do Vorbis

Audio bands in telephony^[1]

Name	Range (Hz)
Narrowband	300–3,400
Wideband	50–7,000
Superwideband	50–14,000
Fullband	20–20,000



<https://opus-codec.org/comparison/>

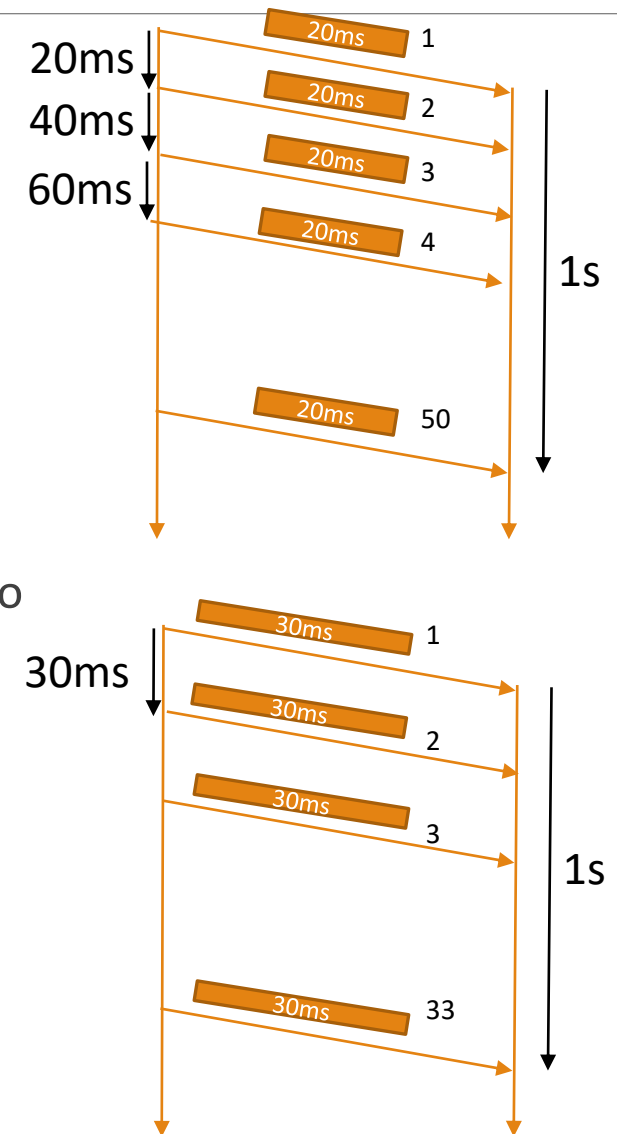
Recomendações ITU-TS para Voz

Recomend.	Técnica de Compressão	Largura de banda da voz (kHz)	Taxa de amostragem (kHz)	Taxa de bit compactado (kbps)	Duração do payload (ms)	Tamanho do payload (bytes)
G.711	PCM não linear	3,4	8	64	20	160
G.722	ADPCM sub-banda	7	16	48, 56, 64		
G.723.1m	MP-MLQ	3,4	8	6.4	30	24
G.723.1a	ACELP	3,4	8	5,3	30	20
G.726	ADPCM	3,4	8	16, 24, 32, 40	15	60
G.728	LD-CELP	3,4	8	16	20	40
G.729A	CS-CELP	3,4	8	8	20	20

Codecs e QUADRO de Voz

Tamanho do quadro de voz (payload)

- Tamanho do payload (em bytes)
 - Taxa do Codec (em bits/sec) x tempo do quadro de voz (s)
- Exemplo 1:
 - Codec G.711 => 64 kbps
 - Tamanho do pacotes = 20ms
 - Tamanho do payload = $(64000 \times 0,02)/8 = 160$ bytes
 - Teríamos $(1000/20) = 50$ pacotes de 160 B de dados a cada segundo
- Exemplo 2:
 - Codec G.711 => 64 kbps
 - Tamanho do pacotes = 30ms
 - Tamanho do payload = $(64000 \times 0,03)/8 = 240$ bytes
 - Teríamos 33 pacotes de 240 B de dados a cada segundo



Codecs e QUADRO de Voz

Problema da sobrecarga de protocolos

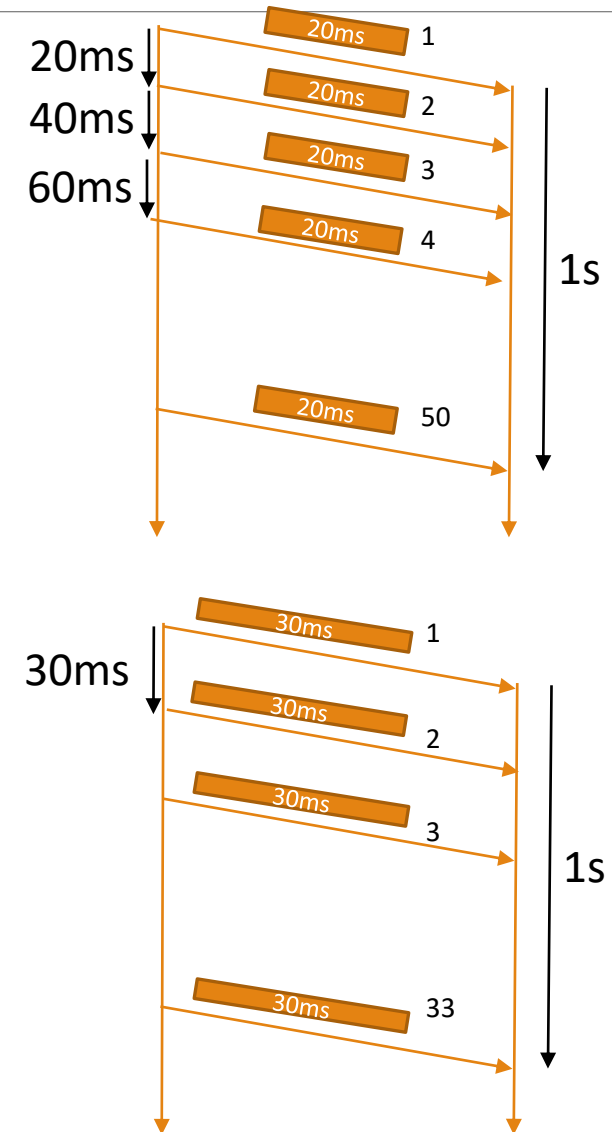
- Para ser transmitido na rede, o quadro de voz deve ser encapsulado em diversos protocolos
 - Até chegar à camada de enlace, aos pacotes de voz vão ser adicionados 40 bytes: RTP (12 bytes) + UDP (8 bytes) + IP (20 bytes) = 40 bytes.
 - Se o quadro é pequeno, a sobrecarga de protocolos é maior:
 - Quadro de voz de 160B: pacote IP será de 40+160=200B, sobrecarga de $40/200 = 20\%$
 - Quadro de voz de 240B: pacote IP será de 280B, sobrecarga de $40/280 = 14\%$



Codecs, Quadros e Pacotes de Voz

Relação entre tamanho de quadro de voz e atraso

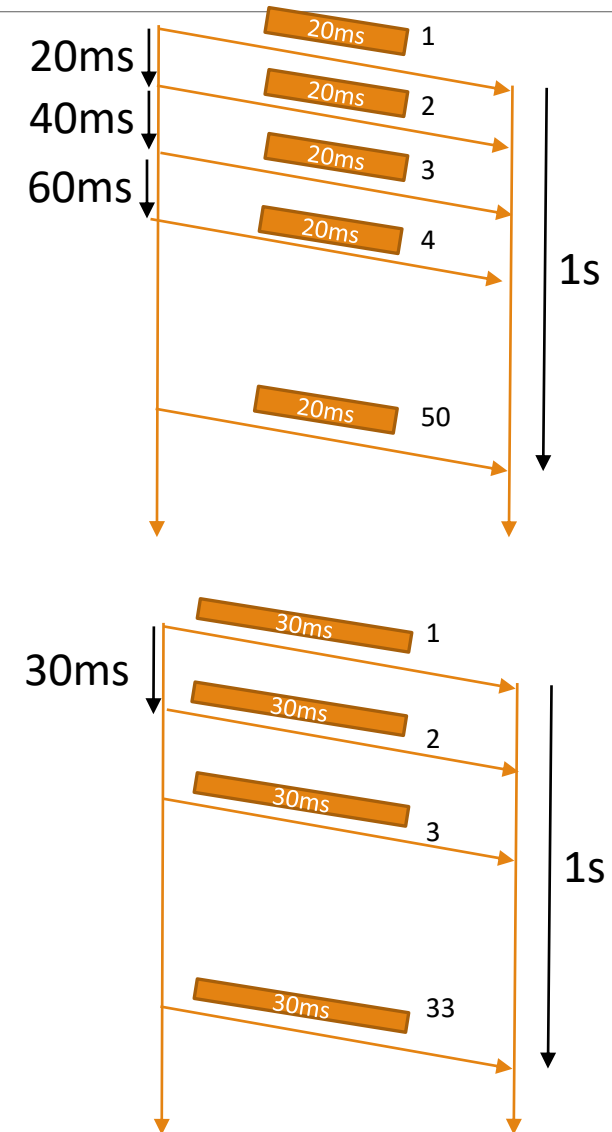
- Para redução do atraso, o codec escolhido deveria ter um quadro de quadro pequeno
- Exemplo 1:
 - Codec G.711 => 64 kbps
 - Tamanho do pacotes = 20ms
 - Tempo de empacotamento será 20ms
- Exemplo 2:
 - Codec G.711 => 64 kbps
 - Tamanho do pacotes = 30ms
 - Tempo de empacotamento será de 30ms



Codecs, Quadros e Pacotes de Voz

Relação entre tamanho de quadro de voz e taxa de transmissão

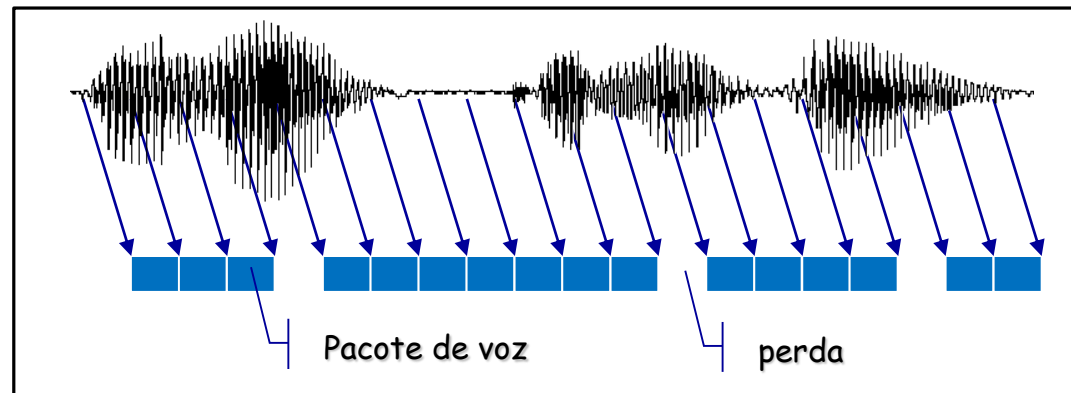
- Quadro pequeno (i.e. menor atraso) gera uma maior taxa de bits devido a sobrecarga dos protocolos
- Exemplo 1: Codec G.711 => 64 kbps
 - Tamanho do pacote de voz = 20ms
 - Tamanho do Pacote IP $40+160 = 200$ B
 - Teríamos 50 pacotes IP de 200 B de dados a cada segundo
 - Taxa de bits é de $50 \cdot 200 \cdot 8 = 80$ kbps
- Exemplo 2: Codec G.711 => 64 kbps
 - Tamanho do pacote de voz = 30ms
 - Tamanho do pacote IP = $40 + 240 = 280$ B
 - Teríamos 33 pacotes de 280 B de dados a cada segundo
 - Taxa de bits é de $33 \cdot 280 \cdot 8 = 73,9$ kbps



Codecs, Quadros e Pacotes de Voz

Relação entre tamanho de quadro de voz e qualidade na ocorrência de perda de pacotes

- Se o pacote de voz é maior, a perda gera um tempo maior de ausência de som na saída
- Maior o pacote, pior a tolerância à perda de pacotes.



Recomendações ITU-TS para Voz

■ G.711

- Usa PCM compandido (escala não linear)
 - Serve para aumentar a resolução de sinais de baixa amplitude
 - Mais importante para os humanos
 - Operando de forma análoga ao ouvido humano
- Dois tipos de escala
 - A-law (Europa e Brasil)
 - μ -law (EUA e Japão)
- Usado na maioria dos backbones telefônicos digitais
- Fluxo de bits de 64 kbps
 - 8 bits por amostra, 8000 amostras/s (uma amostra a cada 125 μ s)
- Supressão de silêncio é opcional
 - Reduz a taxa de bits gerada

Recomend.	Técnica de Compressão	Largura de banda da voz (kHz)	Taxa de amostragem (kHz)	Taxa de bit compactado (kbps)	Duração do payload (ms)	Tamanho do payload (bytes)
G.711	PCM não linear	3,4	8	64	20	160

Recomendações ITU-TS para Voz

Tamanho do Payload

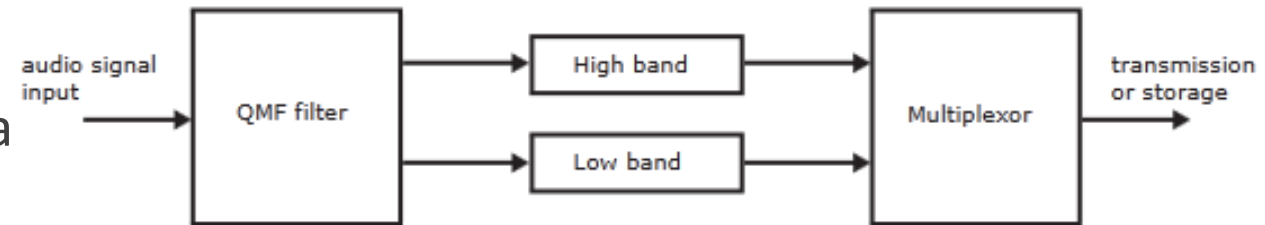
- Quanto menor o payload de voz maior é a sobrecarga dos diversos protocolos de transmissão da voz
- Quanto maior o payload maior é o atraso na aplicação
 - para aguardar a montagem do payload

Recomendação	Taxa de bit compactado (kbps)	Duração do payload (ms)	Tamanho do payload (bytes)
G.711	64	20	160
G.723.1m	6.4	30	24
G.723.1a	5.3	30	20
G.726	16, 24, 32, 40	15	60
G.728	16	20	40
G.729A	8	20	20

Recomendações ITU-TS para Voz

G.722

- Fornece uma melhor qualidade que o G.711 e G.721: Utiliza 14 bits por amostra
- ADPCM Sub-banda: sinal de voz é dividido em duas sub-bandas: alta (4-8kHz) e baixa (0-4kHz), no 64kbps:
 - 2 bits/amostra para banda alta (16 kbps)
 - 6 bits/amostra para banda baixa (48 kbps)
- Próprio para aplicações de videoconferência uma vez que telefones comuns não respondem na faixa de 7kHz



Recomend.	Técnica de Compressão	Largura de banda da voz (kHz)	Taxa de amostragem (kHz)	Taxa de bit compactado (kbps)	Duração do payload (ms)	Tamanho do payload (bytes)
G.711	PCM não linear	3,4	8	64	20	160
G.721	ADPCM	3,4	8	32		
G.722	ADPCM sub-banda	7	16	48, 56, 64		

Recomendações ITU-TS para Voz

G.723.1

- Opera a 6,4 kbps (Multipulse-Maximum Likelihood Quantification) e a 5,3 kbps (Algebraic-Code-Excited Linear Prediction)
- Em cada janela de 30 ms do sinal de voz
 - são analisadas 240 amostras de 16 bits do sinal de voz (tomadas a 8kHz) para identificação de padrões repetitivos (pitches) e são gerados 12 ou 10 códigos de 16 bits, conforme o algoritmo esteja configurado para uma taxa de 6,3 ou 5,3 kbps
- Valor típico de tamanho do pacote de voz (payload) é de 30ms (20 ou 24 bytes)

Recomend.	Técnica de Compressão	Largura de banda da voz (kHz)	Taxa de amostragem (kHz)	Taxa de bit compactado (kbps)	Duração do payload (ms)	Tamanho do payload (bytes)
G.711	PCM não linear	3,4	8	64	20	160
G.723.1m	MP-MLQ	3,4	8	6.4	30	24
G.723.1a	ACELP	3,4	8	5.3	30	20

Recomendações ITU-TS para Voz

G.726

- O G.726 utiliza o ADPCM a 40, 32, 24 e 16 kbps
- Sinal de voz é amostrado a 8kHz, codificado em 8 bits (leis A ou m) e são transmitidas diferenças entre amostras com 5, 4, 3 ou 2 bit em quantificação adaptativa
- Valor típico de tamanho do pacote de voz (payload) é de 15ms (60 bytes)

Recomend.	Técnica de Compressão	Largura de banda da voz (kHz)	Taxa de amostragem (kHz)	Taxa de bit compactado (kbps)	Duração do payload (ms)	Tamanho do payload (bytes)
G.711	PCM ñ linear	3,4	8	64	20	160
G.723.1m	MP-MLQ	3,4	8	6.4	30	24
G.723.1a	ACELP	3,4	8	5.3	30	20
G.726	ADPCM	3,4	8	16, 24, 32, 40	15	60
G.728	LD-CELP	3,4	8	16	20	40
G.729A	CS-CELP	3,4	8	8	20	20

Recomendações ITU-TS para Voz

G.728

- Técnica de codificação LD-CELP (Low-Delay, Code-Excited Linear Prediction), gerando uma taxa de bits de 16 kbps
- Tabela (codebook) utilizada é formada por 1024 (2^{10}) valores
 - contém os valores de códigos (vetores) que representam as possíveis amostras do sinal de voz
- Em cada janela de $625\mu\text{s}$ do sinal de voz são analisadas 5 amostras de 8 bits e é gerado 1 código de 10 bits

Recomend.	Técnica de Compressão	Largura de banda da voz (kHz)	Taxa de amostragem (kHz)	Taxa de bit compactado (kbps)	Duração do payload (ms)	Tamanho do payload (bytes)
G.711	PCM ñ linear	3,4	8	64	20	160
G.723.1m	MP-MLQ	3,4	8	6.4	30	24
G.723.1a	ACELP	3,4	8	5.3	30	20
G.726	ADPCM	3,4	8	16, 24, 32, 40	15	60
G.728	LD-CELP	3,4	8	16	20	40
G.729A	CS-CELP	3,4	8	8	20	20

Recomendações ITU-TS para Voz

G.729

- Bastante popular em aplicações de voz sobre frame relay e em modems V.70 para voz e dados
- G.729 Técnica de codificação LD-CELP gerando uma taxa de bits de 8 kbps e G.729A a codificação CS-ACELP (Algebraic-ACELP)

Recomend.	Técnica de Compressão	Largura de banda da voz (kHz)	Taxa de amostragem (kHz)	Taxa de bit compactado (kbps)	Duração do payload (ms)	Tamanho do payload (bytes)
G.711	PCM ñ linear	3,4	8	64	20	160
G.723.1m	MP-MLQ	3,4	8	6.4	30	24
G.723.1a	ACELP	3,4	8	5.3	30	20
G.726	ADPCM	3,4	8	16, 24, 32, 40	15	60
G.728	LD-CELP	3,4	8	16	20	40
G.729A	CS-CELP	3,4	8	8	20	20

Recomendações ITU-TS para Voz

G.729

- Em cada janela de 10ms do sinal de voz são analisadas 80 amostras de 8 bits para geração de 10 códigos de 8 bits
- Valor típico de tamanho do pacote de voz é de 20ms (20 bytes)

Recomend.	Técnica de Compressão	Largura de banda da voz (kHz)	Taxa de amostragem (kHz)	Taxa de bit compactado (kbps)	Duração do payload (ms)	Tamanho do payload (bytes)
G.711	PCM ñ linear	3,4	8	64	20	160
G.723.1m	MP-MLQ	3,4	8	6.4	30	24
G.723.1a	ACELP	3,4	8	5.3	30	20
G.726	ADPCM	3,4	8	16, 24, 32, 40	15	60
G.728	LD-CELP	3,4	8	16	20	40
G.729A	CS-CELP	3,4	8	8	20	20

Supressão de Silêncio e remoção de sons repetitivos

Compressão da voz via remoção dos períodos de silêncio e de informações redundantes encontradas na fala humana

- Existem informações na fala humana que não são necessárias para que uma comunicação efetiva exista através de uma rede

Sons repetitivos, inerentes à voz, são causados pela vibração das cordas vocais

- transmissão destes sons idênticos não é necessária para efetivação da comunicação e a sua remoção resulta em um aumento de eficiência na utilização da banda de rede

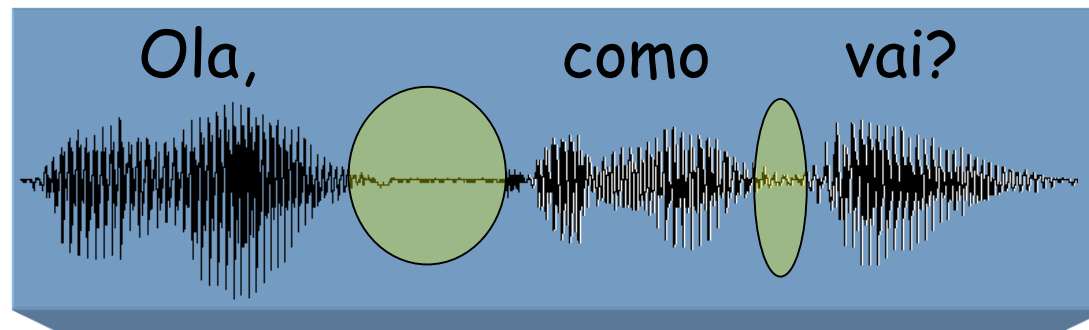
Composição da fala

- 22% do que se fala são componentes essenciais da comunicação
 - devem ser transmitidos para o entendimento do diálogo
- 22% são padrões repetitivos
- 56% representa as pausas entre falas

Supressão do silêncio: Componentes

VAD (Detector de Presença de Voz)

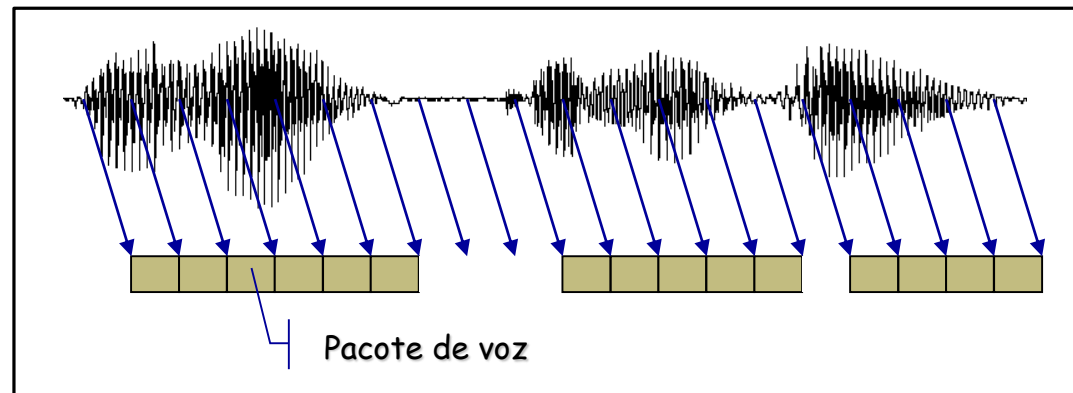
- Responsável por determinar quando o usuário está conversando e quando ele está em silêncio
- É útil para economizar energia no caso de dispositivos que funcionam a bateria
- Deve ser bastante sensível
 - Caso contrário, o início das palavras podem ser perdidas e um silêncio inútil pode ser incluído no final das sentenças
 - Mas ao mesmo tempo não pode ser disparado por ruído de fundo.



Supressão do silêncio: Componentes

DTX (Discontinuous Transmission)

- Capacidade de um codec de parar de transmitir quadros quando o VAD tiver detectado um período de silêncio
- VAD + DTX: modo eficiente de liberar dinamicamente a banda
 - proporcionando uma economia de até 50% da banda
- Alguns codecs avançados não vão interromper a transmissão completamente
 - Em vez disso, vão para um modo de silêncio no qual usam muito menos largura de banda e enviam apenas os parâmetros mínimos para que o receptor possa restituir o ruído de fundo (intensidade, etc.)



Supressão de Silêncio e remoção de sons repetitivos

Alguns pontos devem ser considerados na supressão do silêncio

- Quando a fala é muito frequente, contínua, os ganhos com a supressão do silêncio não são alcançados;
- Como a detecção da presença de voz na transmissão não é imediata
 - Pode ocorrer o corte das primeiras sílabas da locução
 - Fenômeno é denominado de clipping;
- Quando o ruído de fundo é muito alto
 - Torna-se difícil distinguir entre o que é ruído e o que realmente é fala
 - Corre-se o perigo de empacotamento de ruído.

Pontos Importantes

Codecs de Voz

- Entender o que são pacotes de voz e a relação do tamanho do pacote de voz com taxa de bits, atraso e impacto na qualidade quando da perda de pacotes

Supressão de Silêncio

- Entender as vantagens e limitações