# OPUS: O VERDADEIRO CANIVETE SUÍÇO DOS CODECS DE ÁUDIO

*Afonso Eduardo nº68726*
*Pedro Barradas nº 63166*
*Pedro Escudeiro nº67696*

Instituto Superior Técnico
Av. Rovisco Pais, 1049-001 Lisboa, Portugal
E-mail: {afonso.eduardo, pedro.barradas, pedro.n.escudeiro}@ist.utl.pt

**RESUMO**

Este artigo propõe-se a investigar e explicitar as principais características e vantagens do recentemente publicado *codec* de áudio *Opus*, descrevendo-se o seu modo de funcionamento através dos mecanismos que este emprega aquando da codificação e descodificação de sinais de áudio. É também do âmbito deste documento compará-lo qualitativa e quantitativamente com os demais *codecs*. Pretende-se ainda apresentar as suas implicações ao nível do aparecimento de novas aplicações e de alterações tecnológicas e económicas num futuro próximo.

***Palavras-Chave —*** áudio, *codec*, IETF, *Opus*, *streaming*

## 1. INTRODUÇÃO

No decorrer das últimas décadas, tem-se vindo a observar um tremendo avanço ao nível de inúmeras tecnologias e nas mais diversas áreas, sendo que o áudio não é exceção. Com o surgimento de sistemas digitais, ocorreu uma revolução na forma como o som é armazenado e reproduzido, sempre com o objetivo de alcançar mais e melhor com o menor número de recursos possível. Nesse sentido, a indústria mostrou sempre um grande interesse em aprimorar os processos de como se codifica o áudio, nomeadamente aqueles que se encontram associados à sua compressão. Assim, tornou-se, então, necessário aprofundar os conhecimentos sobre o funcionamento do sistema auditivo humano, especialmente as suas limitações. Estes estudos, por sua vez, permitiram a criação de sistemas bastante eficazes que passam pela remoção de partes irrelevantes e redundantes dos sinais de áudio, sendo geralmente agrupados numa categoria apelidada de compressão com perdas.

Naturalmente que este tipo de compressão rapidamente se tornou o predileto de vários prestadores de serviços e, consequentemente, de muitos consumidores. As razões desta adoção são variadas, mas passam sobretudo pelo facto de se conseguir armazenar um maior número de informação útil num dado dispositivo, pela viabilidade que confere às aplicações em tempo real e, não menos importante, pela redução de custos de operação decorrentes da diminuição do tamanho dos ficheiros áudio.

Nos últimos anos, como seria de esperar, a maioria dos codificadores quer de música quer de voz assentaram neste mesmo princípio, sendo o *Advanced Audio Coder* (AAC) e o MPEG-1 *Audio Layer 3* (MP3) referências na indústria da música, tal como o *Adaptive Multi-Rate* (AMR) e o G.729 da ITU-T o são na telefonia. Contudo, estes não são gratuitos, o que significa que qualquer entidade que queira desenvolver aplicações que usem estes formatos não o conseguem fazer sem que seja paga as correspondentes taxas de utilização e de direitos de patente [1]. Ora, esta prática, apesar de já se encontrar bem estabelecida nos mercados *offline*, não foi bem recebida por algumas comunidades *online* uma vez que advogam que tais ações são contrárias ao modelo sobre o qual a Internet assenta, para além de imporem sérias restrições ao nível do desenvolvimento tecnológico e de modelos de negócio alternativos [2]. Neste sentido, surgiram diversos *codecs* de utilização livre dos quais se destacam o Vorbis, para música, e o Speex, para voz. Note-se, porém, que nenhum deles alcançou o grau de sucesso que os seus criadores ambicionavam.

Confiantes que eram capazes de aperfeiçoar o seu trabalho, os criadores do Vorbis e do Speex, membros integrantes da fundação Xiph.Org, continuaram os seus esforços a fim de desenvolver um novo *codec* que integrasse as características das suas soluções anteriores, obtendo, deste modo, um *codec* capaz de atingir a mesma qualidade do Vorbis e com atrasos algorítmicos ainda mais reduzidos que o Speex. Este *codec* foi denominado de CELT e visava, assim, a transmissão de música e voz de alta qualidade.

Entretanto, e numa tentativa de evitar o pagamento de taxas de utilização e direitos de patentes a terceiros, a Skype desenvolve a sua própria solução especificamente vocacionada para a telefonia via Internet, conseguindo codificar voz a débitos extremamente reduzidos [3]. Esta solução foi apelidada de SILK.

Mais tarde, no âmbito da criação e padronização de um sistema de codificação de áudio para aplicações interativas, um protótipo de um formato híbrido, composto pelo SILK e pelo CELT, foi proposto à IETF (*Internet Enginee-*

*ring Task Force*), organização que liderava esta iniciativa. Após ter considerado vários outros *codecs*, a IETF acabou por aceitar este protótipo, referindo que nenhuma outra proposta era tão versátil, nem tão liberal a nível de licenciamento [4]. Apontaram, ainda, que possuía uma qualidade igual ou superior relativamente aos restantes *codecs* então existentes pelo que era inquestionável que se tratava da solução mais favorável em termos técnicos. Note-se, porém, que esta decisão gerou alguma polémica, havendo vários participantes que manifestaram o seu descontentamento face à validação da proposta, especialmente os detentores de patentes de tecnologias concorrentes [5].

No seguimento desta decisão, foi então criado um grupo de trabalho de forma a finalizar o protótipo, tendo assim nascido, meses mais tarde, o *Opus* que promete revolucionar o mundo áudio.

## 2. PRINCIPAIS CARACTERÍSTICAS

O *Opus* caracteriza-se por ser um *codec* muito versátil principalmente devido à sua capacidade de operar desde débitos binários muito baixos (6 kbit/s) até a relativamente elevados (510 kbit/s), mantendo sempre um atraso dentro da gama dos 5 a 65 ms [6]. Tais resultados devem-se a uma característica que lhe é bastante peculiar: possui duas camadas sobre as quais pode transitar suavemente sem introduzir distorção no envio da informação o que, por sua vez, revela ser de grande utilidade dado que cada uma destas camadas se encontra otimizada para operar sobre um certo tipo de conteúdo e para uma certa gama de débitos. A associação destas duas camadas permite com que o *Opus* se adapte tanto às variações do conteúdo que transmite como da própria rede sobre a qual a ligação é efetuada sem necessidade de renegociar a sessão.

Tendo em conta esta mentalidade de escalabilidade, é natural que este *codec* possua parâmetros altamente variáveis e, como tal, tem de suportar um leque suficientemente grande de larguras de banda (LB), o que se reflete em ritmos de amostragem diferentes, e de tramas de áudio de diferente duração (controlo do atraso do sistema). Alguns destes parâmetros encontram-se representados na tabela 1. Note-se, ainda, que, na qualidade de codificador com perdas, o *Opus* apenas tenta preservar os detalhes que são audíveis pelos humanos e uma vez que o ouvido humano não consegue ouvir frequências superiores a 20kHz, tal informação é considerada irrelevante, sendo, portanto, descartada.

| Designação da LB | LB de Áudio | Ritmo de Amostragem |
|---|---|---|
| NB (Banda Estreita) | 4 kHz | 8 kHz |
| MB (Banda Média) | 6 kHz | 12 kHz |
| WB (Banda Larga) | 8 kHz | 16 kHz |
| SWB (Banda Super Larga) | 12 kHz | 24 kHz |
| FB (Banda Completa) | 20 kHz | 48 kHz |

**Tabela 1. Larguras de banda suportadas pelo *Opus* [6]**

### 2.1. Camadas e Modos de Funcionamento

Tal como referido anteriormente, o *Opus* possui duas camadas: SILK e CELT. A primeira, tendo sido baseada num *codec* desenvolvido pela Skype, serve especificamente para comunicações voz a baixo débito com a condicionante de não ser apropriada para música. Já a segunda, que tenta corrigir esta limitação, codifica música e voz para débitos superiores.

À semelhança de muitos outros sistemas onde o todo é mais do que a soma das suas partes, o *Opus*, para um dado instante, também não se limita a explorar as características intrínsecas de apenas uma camada. Neste sentido, surge um modo de funcionamento adicional que tenta tirar o máximo partido decorrente da sinergia entre o SILK e o CELT, aproveitando a alta compressão de sinais de voz oferecida pelo primeiro e a eficiência de codificação obtida pelo uso do CELT para altas frequências.

É, ainda, de realçar o facto de que o *Opus* é capaz de alternar entre modos de funcionamento no decurso de uma sessão.

| Modo | LB | Duração das Tramas[ms] | Aplicação | *Bitrate* Suportada |
|---|---|---|---|---|
| SILK | NB | 10; 20; 40; 60 | Comunicação voz a baixo débito (< 30kbit/s) | Variável (VBR); Constante (CBR) |
| | MB | | | |
| | WB | | | |
| Híbrido | SWB | 10, 20 | Comunicação voz de alta qualidade | |
| | FB | | | |
| CELT | NB | 2,5; 5; 10; 20 | Comunicação voz com baixa latência; Música | |
| | WB | | | |
| | SWB | | | |
| | FB | | | |

**Tabela 2. Modos de funcionamento do *Opus* [6]**

### 2.2. Parâmetros de Controlo

Com vista à adaptabilidade, o *Opus* possui um conjunto de parâmetros que se altera dinamicamente durante o seu funcionamento sem que seja necessário interromper a transmissão:

- ***Bitrate***: suporta débitos desde os 6kb/s até aos 512kb/s com incrementos de 0,4kb/s o que corresponde a mais de 1200 combinações possíveis.
- **Número de canais**: capaz de transmitir tanto tramas monofónicas como estereofónicas num único fluxo de dados. A cada instante, o codificador tenta tomar a melhor decisão com base no débito da ligação.
- **Duração de cada trama**: para além da possibilidade de existirem tramas de duração diferente, o *Opus* tem a capacidade de aglomerá-las em pacotes de até 120ms. O aumento da duração das tramas conduz a uma codificação mais eficiente, mas tam-

bém a um atraso maior pelo que este mecanismo de controlo é de extrema importância.

- **Largura de banda de áudio**: o codificador escolhe qual é a LB que conduz a uma melhor *performance* tendo em conta o débito binário disponível.
- **Complexidade**: permite ajustar a sua exigência em termos de CPU face à qualidade/*bitrate*.
- **Resistência à perda de pacotes**: adapta a sua capacidade de correção com base na *bitrate* disponível.
    - Ajusta o grau com que explora a correlação entre tramas. Esta correlação permite enviar a mesma quantidade de informação a débitos inferiores à custa da propagação de erros.
    - Possui um mecanismo que reforça a sua robustez, determinando, numa primeira etapa, quais são os pacotes enviados que possuem informações críticas e, em seguida, retransmite-os a um débito inferior.
- **Transmissão descontínua (DTX)**: reduz a *bitrate* quando o sinal a enviar é silencioso ou ruído de fundo, transmitindo apenas uma trama a cada 400ms.

## 3. ARQUITETURA

O diagrama de blocos do funcionamento do *Opus* apresenta-se na figura 1.

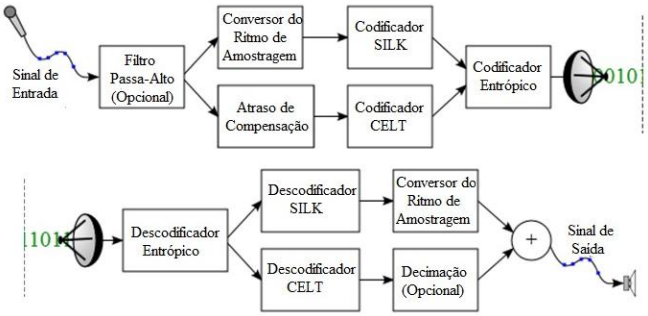


**Fig. 1. Esquema do funcionamento do *Opus* [3]**

À entrada do sistema encontra-se um filtro passa-alto com uma frequência de corte de 50Hz e cuja função é remover as componentes mais baixas do espectro que contêm ruído de fundo. No entanto, para aplicações que envolvam música, torna-se ainda necessário adicionar um detetor que permita reduzir o valor da frequência de corte de forma a não degradar o sinal de entrada.

Seguidamente, e dependendo do modo de funcionamento, o sinal filtrado é transmitido aos codificadores SILK e/ou CELT. Nos modos SILK e híbrido, o sinal sofre uma conversão no ritmo de amostragem para que seja compatível com o funcionamento interno do codificador SILK. Neste último modo, é ainda necessário a introdução de um atraso à entrada do codificador CELT por motivos de sincronismo.

Após esta etapa, o sinal resultante é enviado para um codificador entrópico, algoritmo que se baseia na codificação aritmética (*range encoding*), para posterior transmissão. É importante assinalar que o fluxo de *bits* decorrente desta operação obedece a uma estrutura definida pela norma RFC 6716 de forma a assegurar a interoperabilidade entre sistemas provenientes de diferentes fabricantes.

No lado da receção, segue-se um conjunto de operações inversas às anteriormente descritas que permitem obter um sinal de áudio que se assemelha ao original, sendo esta relação tanto mais próxima quanto maior for a largura de banda definida. Mais uma vez, por motivos de interoperabilidade, existe um módulo à saída do descodificador CELT responsável pela redução do ritmo de amostragem para um suportado pelo *hardware* de áudio à saída.

Ainda neste seguimento, é importante referir que apenas o descodificador é normativo, existindo, assim, liberdade para se realizar modificações na implementação do codificador, facto este que potencia o desenvolvimento e competitividade tecnológica.

### 3.1. Camada CELT

A camada CELT, *Constrained Energy Lapped Transform*, baseia-se num modelo psicoacústico de conservação de energia e aplica conceitos da codificação por transformação [7]. Como tal, faz uso da Transformada de Cosseno Discreta Modificada (MDCT) à semelhança de outros *codecs*, tais como o MP3 ou o AAC. Esta transformação assenta no facto de que qualquer sinal pode ser representado como a soma de ondas sinusoidais (cossenos) de diferentes frequências e que a cada uma destas sinusoides se encontra associado um determinado peso (energia). Desta forma, para representar um sinal analógico basta identificar os pesos e as correspondentes frequências que o constituem. A MDCT é, assim, uma ferramenta de análise uma vez que é capaz de determinar estes parâmetros o que, por sua vez, permite reduzir significativamente a quantidade de informação que é necessária para descrever um dado sinal.

É, também, de referir que o termo "modificada" da MDCT decorre do facto de esta usar janelas contínuas que se sobrepõem a tramas adjacentes, ao contrário da DCT que codifica as tramas como blocos independentes, não havendo, neste último caso, uma transição suave entre estes pelo que o sinal fica percetivamente pior (efeito de bloco).

Tal como todos os sistemas, este também possui limitações e a sua diz respeito à incapacidade de conseguir obter uma boa resolução no tempo e na frequência simultaneamente. Assim, caso se queira obter, no lado da receção, um sinal com o menor atraso possível, isto é, uma boa resolução temporal torna-se necessário que o mesmo seja dividido em pequenas tramas durante a sua análise e codificação o que, por sua vez, conduz a uma degradação da resolução na frequência e, eventualmente, ao espalhamento espectral. Uma outra implicação que daqui decorre é a necessidade de

se ter que recorrer a mais informação adicional para representar o mesmo sinal, implicando um aumento do débito caso se queira manter a mesma qualidade. Neste âmbito, o CELT possui uma série de características que tentam combater estes efeitos as quais são apresentadas ao longo das secções seguintes.

Antes de se apresentar o codificador e o seu descodificador, convém ainda referir uma outra característica que é tão importante ao ponto de estar na génese do nome CELT. Tal como referido anteriormente, o CELT baseia-se num modelo de conservação de energia e a razão para tal escolha advém do facto de se obter uma melhor qualidade se o objetivo for a preservação a energia das bandas constituintes do sinal original ao invés da minimização do seu erro quadrático médio [8]. Com este intuito, foi desenvolvido uma técnica que consiste em determinar e codificar em separado a energia do sinal, garantindo que o valor da mesma é codificado com uma resolução suficientemente alta. Esta técnica apesar de ser bastante simples, revela-se de extrema utilidade especialmente a baixos débitos onde se torna necessário quantificar os coeficientes da MDCT com poucos níveis.

### *3.1.1. Codificador*

O esquema do codificador desta camada apresenta-se na figura 2.

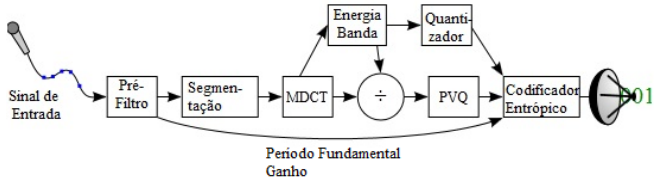


**Fig. 2. Esquema do codificador CELT [7]**

De acordo com este esquema, constata-se que o sinal passa por um pré-filtro que é responsável pela sua análise, bem como pela identificação das suas harmónicas principais através de um preditor de tom. Este processo visa melhorar a qualidade de sinais com conteúdo periódico.

Em seguida, o sinal resultante (com as harmónicas principais atenuadas) é dividido em tramas parcialmente sobrepostas as quais são, posteriormente, representadas em coeficientes no domínio da frequência através da aplicação da MDCT. Adicionalmente, prevê-se que a trama original possa ser subdividida em tramas de ainda menor duração com o intuito de aumentar a resolução temporal e, consequentemente, reduzir o pré-eco. No entanto, este procedimento pode conduzir a que, numa ou mais MDCT's, os coeficientes para algumas bandas sejam nulos, o que origina artefactos desagradáveis pois não permite conservar a energia do sinal. De forma a impedir este efeito, para cada banda onde tal é detetado, é introduzido um sinal pseudoaleatório. Note-se que a energia deste sinal não é importante uma vez que o mesmo será normalizado à energia do sinal original, como é descrito adiante.

Seguidamente, os coeficientes são agrupados em bandas, que se assemelham às faixas críticas da audição humana, sendo a energia de cada uma destas determinada e utilizada para os normalizar. Este procedimento de separação em bandas permite explorar a irrelevância do sinal, uma vez que o ouvido humano possui uma resolução espectral diferencial e limitada, permitindo, assim, representar com menos *bits* os coeficientes correspondentes às bandas onde o ouvido é menos sensível. É de referir que a energia de cada banda é, ainda, comprimida e quantificada para posterior transmissão. Esta compressão baseia-se no facto do descodificador ser capaz de prever o seu valor a partir da trama anterior, bem como da energia da banda que a antecede pelo que o codificador apenas necessita de enviar o erro decorrente desta previsão.

Após a obtenção dos coeficientes normalizados, que mais não são que descritores da forma da energia do sinal, segue-se a etapa da sua quantização que faz uso de um tipo de quantização vetorial, a PVQ (Quantização Vetorial Piramidal) [9]. Esta quantificação é muito mais eficiente do que a quantização escalar, pois possui a particularidade de conseguir aproximar uma distribuição multidimensional com recurso a um número finito de vetores pertencentes a esse espaço. Desta forma, esta fase consiste em agrupar os coeficientes normalizados num vetor, caso pertençam à mesma banda, e, em seguida, procura no dicionário (livro de código) da PVQ pelo vetor mais semelhante. Finalmente, os índices (posições dos vetores no dicionário) referentes às bandas que constituem cada trama são codificados entropicamente.

### *3.1.2. Descodificador*

No que diz respeito à descodificação, muitas das operações descritas anteriormente são revertidas. Assim, inicialmente, são descodificadas as componentes (energia e coeficientes normalizados de cada banda) que permitem reconstruir o espectro de uma trama. A representação temporal de cada trama é, então, obtida através da aplicação da MDCT, sendo estas interligadas através de um processo denominado WOLA (*Weighted Overlap Add*). As harmónicas principais são amplificadas aquando da passagem do sinal resultante pelo pós-filtro e à sua saída, obtém-se o sinal descodificado.

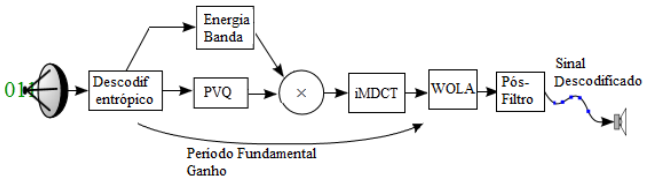


**Fig. 3. Esquema do descodificador CELT [7]**

### *3.1.3. Características adicionais*

A camada CELT, para além de conseguir uma resolução temporal variável através da manipulação da duração das tramas, é também capaz de ajustar a resolução temporal ou frequencial de qualquer banda constituinte de uma dada

trama o que permite uma menor degradação da qualidade do sinal aquando da ocorrência de transientes.

Uma outra característica importante diz respeito à capacidade de codificação a baixo débito de sinais que apresentam um espectro difuso. Nesta situação, existem demasiados coeficientes com valores semelhantes, não sendo, contudo, possível a transmissão da sua grande maioria o que origina o aparecimento de sons metálicos aquando da descodificação. De forma a impedir que tal suceda, é realizada uma transformação do sinal no lado do emissor que posteriormente é revertida no lado do recetor.

Por último, mas de igual relevância, o CELT usa uma técnica (*band folding*) que permite reutilizar os coeficientes normalizados correspondentes a bandas de menor frequência nas de maior frequência. Este procedimento é bastante útil quando se opera a baixos débitos, não sendo possível a transmissão de todos os coeficientes. Como a energia de cada banda é sempre transmitida, o CELT consegue, mesmo assim, manter a energia do sinal o que conduz a uma melhoria na perceção do mesmo.

### 3.2. Camada SILK

A camada SILK do Opus trata-se de uma ferramenta especialmente otimizada para a codificação de sinais de voz que tenta explorar ao máximo a sua redundância, pois apresentam um certo grau de previsibilidade, bem como a sua irrelevância, dado que o ouvido humano possui uma sensibilidade limitada. Com este intuito, o algoritmo descreve estes sinais através de uma representação paramétrica, simplificando-os sem que tal seja percetível ao ouvido humano.

#### *3.2.1. Modelo*

O modelo adotado pela camada SILK tenta imitar a forma como a voz é produzida pelo aparelho fonador humano, isto é, tenta modelar o processo da emissão de ar pelos pulmões e a posterior passagem do mesmo pela glote e pelo trato vocal [10]. Deste modo, o ar emitido pelos pulmões pode ser representado, em linhas gerais, por um gerador de ruído branco (excitação); já as vibrações das cordas vocais e as ressonâncias do trato vocal apresentam-se como filtros que transformam o sinal aleatório num sinal correlacionado. É, também, de referir que é possível dividir a voz em segmentos de pequena duração (tramas), aproximadamente de 20ms. As tramas, por sua vez, podem ser classificadas consoante dois tipos de som:

- **Sons vozeados** (ou sonoros) são produzidos pela vibração regular das cordas vocais e, por esta razão, apresentam uma forma de onda quase periódica com um determinado período fundamental. Naturalmente, esta periodicidade tende a alterar-se ao longo da duração de segmentos de fala uma vez que a tensão aplicada às cordas vocais varia.
- **Sons não vozeados** (ou surdos) não fazem uso das cordas vocais e resultam num sinal pouco correlacionado, tendo a correlação resultante sido introduzida apenas pelas ressonâncias do trato vocal.

Durante o processo da fala, quer para sons vozeados como para surdos, verifica-se que a forma das cavidades que constituem o trato vocal tomam diferentes formas, porém a sua variação no tempo é suficientemente lenta de tal modo que num período de uma trama se pode considerar que as suas características se mantêm aproximadamente constantes [11]. Assim, para modelar o trato vocal, o SILK faz uso de um filtro cujos coeficientes variam de trama para trama. Acresce, ainda, o facto de considerar que uma trama pode ser representada a partir da combinação linear de tramas imediatamente anteriores, pelo que o modelo é comumente conhecido como codificação por predição linear (LPC). Note-se, finalmente, que os coeficientes do filtro mais não são do que os pesos usados para esta combinação linear.

À semelhança do caso anterior, também é possível definir segmentos temporais nos quais a glote pode ser modelada por parâmetros fixos; a sua duração é, porém, de menor duração (5ms) pelo que se denominam de subtramas. Assim, para cada subtrama, é possível extrair um período fundamental que está associado à vibração das cordas vocais e um conjunto de coeficientes. Estes dois parâmetros são de extrema importância dado que, em vez de se explorar a correlação entre amostras consecutivas, a correlação é, agora, explorada entre amostras espaçadas de múltiplos do período fundamental. Este facto leva a que se faça distinção do caso anterior, pelo que estes coeficientes são denominados de coeficientes de predição de longa duração (LTP).

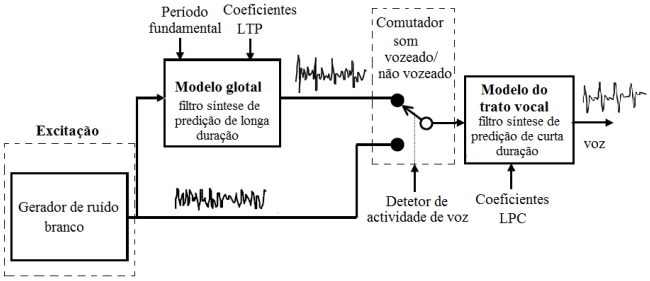


**Fig. 4. Modelo do SILK [10], [12]**

#### *3.2.1. Codificador*

Tendo em conta o referido na secção anterior, rapidamente se percebe que a função do codificador apenas corresponde à determinação e envio do conjunto de parâmetros que melhor descreve o sinal de voz à entrada do sistema para um dado instante.

Para tal, inicialmente, o sinal passa pelo detetor de atividade de voz que auxilia o bloco seguinte a decidir se as tramas são vozeadas ou não vozeadas, sendo essa classificação feita com base no conhecimento que os sons vozeados possuem uma maior energia, especialmente nas baixas frequências [11]. Assim, numa primeira fase, o sinal de cada trama é separado em quatro sub-bandas e é calculada a energia presente em cada uma. De seguida, é determinada a relação sinal-ruído das mesmas, bem como a média resultante, conseguindo, a partir daqui, estimar o nível da atividade da voz e o declive espetral de cada trama constituinte do sinal.

Segue-se, então, a análise do período fundamental. Neste bloco, para além de se conseguir identificar as tramas vozeadas e não vozeadas, é ainda, para o primeiro caso, determinado o período fundamental de cada uma das suas subtramas, bem como um fator que os correlaciona, indicando, deste modo, a periodicidade do sinal.

A seguinte etapa é uma das mais exigentes, pois diz respeito à análise da predição, isto é, à estimação dos coeficientes de predição. Esta fase apresenta, à semelhança do bloco anterior, um tratamento diferencial consoante a classificação da trama. Deste modo, caso a trama seja vozeada, é realizada a estimação dos coeficientes de longa duração os quais são, em seguida, quantificados. Após a sua quantização, os mesmos, para além de serem enviados para o codificador entrópico, são novamente revertidos e usados para filtrar o sinal o que corresponde a retirar ao som vozeado as componentes periódicas introduzidas pelas cordas vocais. Note-se que é de extrema importância que se use os coeficientes após a sua quantificação uma vez que se pretende que o codificador mantenha o sincronismo com o descodificador. O sinal daqui resultante, conhecido por resíduo da predição de longa duração, é, então, usado para a estimação dos coeficientes de curta duração os quais, à semelhança dos outros, são quantificados e transmitidos, com a condicionante de antes serem convertidos numa representação alternativa denominada de LSF (*Line Spectral Frequency*). Esta transformação dos coeficientes LPC em LSF deve-se apenas ao facto destes últimos apresentarem propriedades desejáveis para quantificação e transmissão, pois tomam uma gama limitada de valores e possuem uma grande correlação intertrama [13]. Por outro lado, caso a trama seja classificada como sendo não vozeada, não existe a necessidade de ser feita a análise LTP e, como tal, são estimados diretamente os coeficientes LPC pelo mesmo método acima referido.

Para além dos processos descritos anteriormente, é de referir que em paralelo com estes encontra-se um módulo que é responsável pela análise do ruído, isto é, o codificador possui um esquema de mascaramento auditivo do ruído de quantificação e, como tal, enaltece-o nas regiões onde o sinal possui uma maior energia e minimiza-o nas de menor energia, que são perceptualmente mais sensíveis [11]. A sua função é, pois, otimizar os parâmetros de controlo do pré-filtro e do quantificador a fim de reduzir a perceção do ruído aquando da descodificação do sinal.

Finalmente, e não menos importante, é à saída do filtro branqueador que se obtém o resíduo de predição, ou excitação, o qual idealmente corresponde ao sinal original completamente descorrelacionado, obtido através da sua filtragem pelos coeficientes que modelam o modelo glotal (LTP) e subsequentemente pelos do modelo do trato vocal (LPC). No entanto, tal não acontece, sendo a excitação também composta por erros devidos à quantificação dos coeficientes e a próprias imprecisões dos modelos usados para caracterizar a glote e o trato vocal.

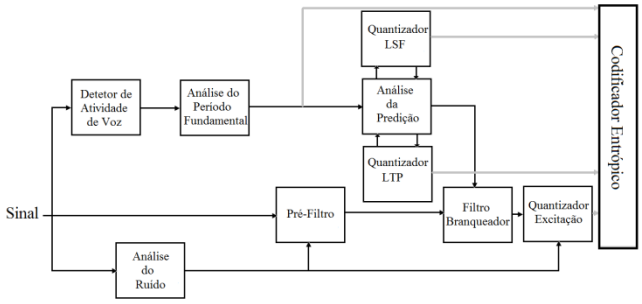


**Fig. 5. Esquema do codificador SILK simplificado [6]**

### *3.2.1. Descodificador*

O processo de descodificação é relativamente simples. Em primeiro lugar, a informação passa pelo descodificador entrópico o que permite obter os parâmetros caracterizadores do sinal, nomeadamente a classificação da trama, a excitação, os coeficientes LPC e, ainda, o período fundamental e os coeficientes LTP caso se trate de uma trama vozeada. A etapa seguinte corresponde a adicionar à excitação a correlação que esta previamente possuía: curta duração para o caso de tramas não vozeadas; longa e curta duração para as vozeadas. No final destes procedimentos, obtém-se o sinal descodificado.

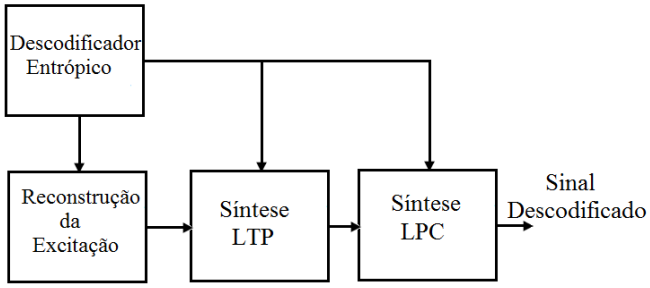


**Fig. 6. Esquema do descodificador SILK simplificado [6]**

## 4. COMPARAÇÃO

Existem vários parâmetros que caracterizam os *codecs* de áudio tais como o ritmo de amostragem, o débito, o atraso e a qualidade. A figura 7 permite comparar o *Opus* com outros *codecs* ao nível destas três primeiras características, observando-se que apesar da sua elevada versatilidade, esta não constitui um fator impeditivo no seu desempenho, dominando os restantes *codecs* ao nível de um baixo atraso.

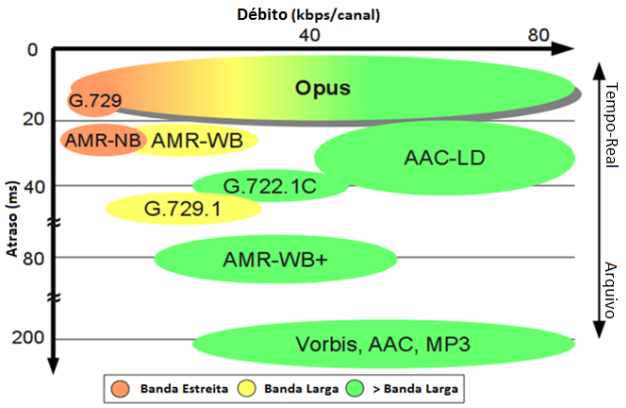


**Fig. 7. Comparação do atraso dos *codecs* a diferentes débitos [3]**

No que diz respeito à qualidade, e atendendo à figura 8, constata-se que o *Opus* revela ser uma alternativa bastante tentadora aquando da sua operação a débitos mais elevados, sendo, inclusive, capaz de destronar as grandes referências a nível da codificação de música. Para *bitrates* intermédias, características de transmissões de voz de alta qualidade, a diferença do mesmo aos restantes é ainda mais notória, constituindo, por isso, a melhor solução para aplicações que façam uso desta banda. É a baixos débitos que o *Opus* tem dificuldade em se manter na vanguarda, uma vez que as extensões AMR conseguem atingir uma qualidade ligeiramente superior, note-se, ainda assim, que este *codec* é claramente superior quando comparado com as alternativas isentas de taxas de utilização.

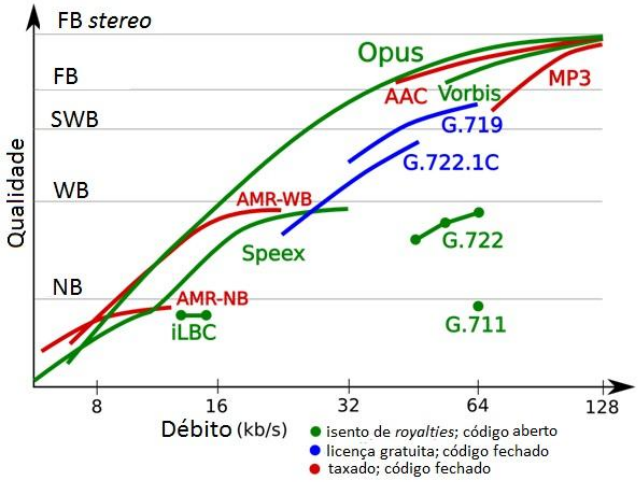


**Fig. 8. Comparação dos *codecs* ao nível da sua qualidade [3]**

## 5. APLICAÇÕES

O *Opus*, tendo sido desenvolvido especialmente para transmissões via Internet, exibe uma grande versatilidade e como tal é ideal para inúmeras aplicações desde comunicações em tempo real, *streaming* ou mesmo o armazenamento.

Assim, este *codec* apresenta um grande potencial para melhorar a qualidade do áudio disponível *online*, algo cada vez mais relevante já que os serviços, quer de comunicação quer de entretenimento, se encontram em expansão para o ciberespaço. Neste sentido, destacam-se não só as aplicações de voz sobre IP, como também videoconferências, telepresença e rádio digital sem contar com os *podcasts* e livros áudio. Surge, ainda, como resultado do baixo atraso que lhe é característico, a possibilidade de realização de atuações musicais ao vivo pela Internet, ou seja torna-se exequível que dois ou mais músicos atuem em conjunto sem que estejam na mesma localização, algo inédito até ao momento.

Por outro lado, uma vez que este *codec* não requer muita informação no início de cada trama (cabeçalho), o mesmo torna-se indicado para codificar ficheiros de pequenas dimensões nomeadamente sons de jogos. Esta reduzida informação inicial é ainda conducente a uma minimização da latência inicial em aplicações ao mesmo tempo que torna a operação de *streaming* menos penosa para os servidores, pois estes não necessitam de guardar informação extra para enviar a clientes que se liguem tardiamente, podendo para os mesmos apenas transmitir um pequeno cabeçalho construído aquando da sua ligação.

Não menos importante, o *Opus* constitui um elemento fulcral no futuro WebRTC , diminutivo para chat em tempo real, o qual permitirá aos *browsers*, como o Chrome e o Firefox, suportar transmissões de voz sobre IP, tarefas atualmente realizadas com recurso a software dedicado como o Skype.

É, ainda, de referir que o *Opus*, apesar do seu processo de padronização ter sido concluído à relativamente pouco tempo, já se encontra presente em inúmeras aplicações o que, mais uma vez, releva o papel preponderante que o mesmo terá num futuro próximo. A nível de sistemas VoIP, destacam-se as aplicações como o Mumble, o TeamSpeak e o Line2. Adicionalmente, está previsto que o Skype também passe a suportar este *codec* numa próxima versão. Relativamente ao *streaming* de áudio, o mesmo já se encontra integrado no Icecast, o que é natural uma vez que este também foi desenvolvido pela Xiph.Org. Finalmente, ao nível de programas de reprodução de conteúdos audiovisuais, tanto o VLC como o foobar2000 já o adotaram.

## 6. LICENCIAMENTO

O *Opus* surge do facto de existirem demasiados *codecs* de áudio no mercado. Tal situação é explicada, em grande parte, pelas restrições de licenciamento que um elevado número deles exibe, uma vez que muitas das empresas ainda apresentam uma inércia considerável em alicerçar o seu negócio em tecnologias pertencentes à concorrência. É, neste sentido, que o *Opus* pretende ser uma ferramenta unificadora, conseguindo aliar à qualidade, um licenciamento flexível. Assim, os criadores do Opus, para além de livremente disponibilizarem a sua especificação, também fornecem a sua implementação a nível de codificador e descodificador, os quais, por sua vez, podem ser integrados, modificados e usados em qualquer aplicação sem nenhum custo

[14]. Este *codec*, não se encontra, portanto, associado a qualquer sistema proprietário, no entanto, tendo resultado de uma colaboração entre a Broadcom, a Microsoft (através da Skype) e a Xiph.Org, as mesmas detêm patentes que servem principalmente para defender os interesses dos seus utilizadores de eventuais ataques judiciais provenientes de outras empresas.

Apesar das entidades que participaram ativamente no desenvolvimento do *Opus* terem declarado que as suas patentes são compatíveis com os princípios de código aberto e de livre acesso e isentas de taxas de utilização, aquando da padronização do Opus pela IETF, outras empresas, nomeadamente a Huawei e a Qualcomm, também declararam que as mesmas detinham direitos intelectuais sobre este *codec*, mas, ao contrário das primeiras, estas exigiam que as suas patentes fossem sujeitas a taxas. Seguiu-se, então, uma fase de investigação por parte da Xiph.Org no sentido de provar que estas alegações não tinham fundamento. Tal foi conseguido, mas à custa de um investimento significativo de recursos humanos o que, por sua vez, abrandou o processo de desenvolvimento do *Opus*. Mais tarde, foi revelado que tanto a Huawei como a Qualcomm não tinham realizado nenhum estudo prévio que validasse as suas declarações, tratando-se apenas de meras suposições o que, apesar de eticamente incorreto, é considerado legal e, como tal, estas ações não tiveram repercussões negativas para nenhuma destas empresas.

## 7. CONCLUSÃO

O *Opus* assenta numa tecnologia de facto surpreendente: apresenta uma melhor qualidade que a esmagadora maioria dos *codecs* existentes no mercado, garantindo, simultaneamente, um atraso significativamente menor. Adicionalmente, é expectável que estas vantagens possam vir a tornar-se ainda mais notórias, considerando o possível aperfeiçoamento do codificador de referência por parte de terceiros. O *Opus* destaca-se, também, por ser extremamente versátil e, como tal, é capaz de operar em múltiplos cenários, adaptando-se especialmente bem ao ambiente *online*. É, por isso, ideal para aplicações VoIP ou de telepresença.

Apesar de ter sido recentemente padronizado pela IETF, já existe uma panóplia de aplicações que o suportam ou que se prepararam para a sua integração, encontrando-se igualmente presente em vários estudos de comunicações móveis. Com o futuro encerramento da rede GSM, estas comunicações serão forçadas a recorrer a sistemas de comutação de pacotes. Tendo o *Opus* sido especialmente desenvolvido para comunicações sobre IP, o mesmo contempla todos os mecanismos necessários para uma transmissão eficiente e fiável de informação, constituindo, naturalmente, uma excelente opção no futuro das tecnologias móveis.
Em suma, o *Opus* terá certamente um papel preponderante no futuro das mais diversas áreas dos sistemas de comunicação audiovisual quer sejam estas de comunicação, reprodução ou gravação de áudio.

## 8. REFERÊNCIAS


[1] Wikipédia, “Comparison of audio formats,” 18 Dezembro 2012. [Online]. Disponível: http://en.wikipedia.org/wiki/Comparison_of_audio_formats. [Acedido em Dezembro 2012].

[2] Xiph.Org, “OpusFAQ,” 19 Outubro 2012. [Online]. Disponível: https://wiki.xiph.org/OpusFAQ. [Acedido em Dezembro 2012].

[3] Wikipédia, “Opus (audio format),” 20 Dezembro 2012. [Online]. Disponível:http://en.wikipedia.org/wiki/Opus_%28audio_format%29. [Acedido em Dezembro 2012].

[4] S. Shankland, “How corporate bickering hobbled better Web audio,” CNET, 17 Agosto 2012. [Online]. Disponível: http://news.cnet.com/8301-1023_3-57494622-93/how-corporate-bickering-hobbled-better-web-audio/. [Acedido em Novembro 2012].

[5] J. D. Rosenberg, “Internet Wideband Audio Codec (codec),” IETF, 21 Maio 2012. [Online]. Disponível: http://datatracker.ietf.org/wg/codec/management/shepherds/draft-ietf-codec-opus/writeup/. [Acedido em Novembro 2012].

[6] J. Valin, K. Vos e T. Terriberry, “Definition of the Opus Audio Codec,” IETF, Setembro 2012. [Online]. Disponível: http://tools.ietf.org/html/rfc6716. [Acedido em Novembro 2012].

[7] Wikipédia, “CELT,” 31 Outubro 2012. [Online]. Disponível: http://en.wikipedia.org/wiki/CELT. [Acedido em Novembro 2012].

[8] Xiph.Org, “Next generation audio: CELT,” 23 Dezembro 2010. [Online]. Disponível: http://people.xiph.org/~xiphmont/demo/celt/demo.html. [Acedido em Dezembro 2012].

[9] T. Fischer , "A pyramid vector quantizer," Information Theory, IEEE Transactions on , vol.32, no.4, pp. 568- 583, Julho 1986.

[10] W. C. Chu, Speech Coding Algorithms: Foundation and evolution of standardized coders, San Jose, CA: Wiley, 2004.

[11] C. M. Ribeiro, “Processamento Digital de Fala,” Maio 2012. [Online]. Disponível: http://www.deetc.isel.ipl.pt/comunicacoesep/disciplinas/pdf/index.html. [Acedido em Dezembro 2012].

[12] S. V. Vaseghi, Advanced Digital Signal Processing and Noise Reduction, Wiley, 2000.

[13] L. Hanzo, F. C. Somerville e J. Woodard, Voice and Audio Compression for Wireless Communications, Chichester, West Sussex, Inglaterra: Wiley, 2007.

[14] Xiph.Org, “Opus licensing,” Dezembro 2012. [Online]. Disponível: http://www.opus-codec.org/license/. [Acedido em Dezembro 2012].