



Artificial Intelligence &

Machine Learning





Audição Cognitiva

Prof . André Silva de Carvalho Email: profandre.carvalho@fiap.com.br

www.linkedin.com/in/andresilvadecarvalhohttp://lattes.cnpq.br/6876528572507972

2020

Artificial Intelligence & Machine Learning





Artificial Intelligence & Machine Learning FIAP



Para próxima aula

Segunda atividade



Gravar três áudios com a resposta para pergunta:

"Você acredita que a redes sociais digitais ajudam a aumentar as posições extremistas? Por que você tem esta sua opinião?."

Por gentileza, tragam fones de ouvido.

Artificial Intelligence & Machine Learning FIAP



Conceitos iniciais

Uma reflexão inicial



1997



2011



ALPHAGO

2016

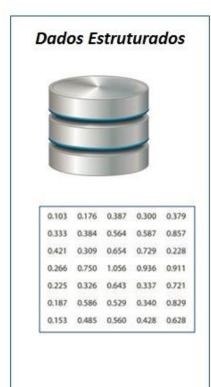


Os dados que existem nas empresas



- Estima-se que 80% dos dados atuais das empresas são do tipo não estruturados (December 2016 (15:4) | MIS Quarterly Executive.)
- Com o advento da internet e as tecnologias que permitem a coleta e armazenamento de dados, permitiu um aumento do volume de dados, que gerou um ruído no processo de análise

(Problems of Economic Transition, vol. 57, no. 12, 2015)





Transformar som em texto



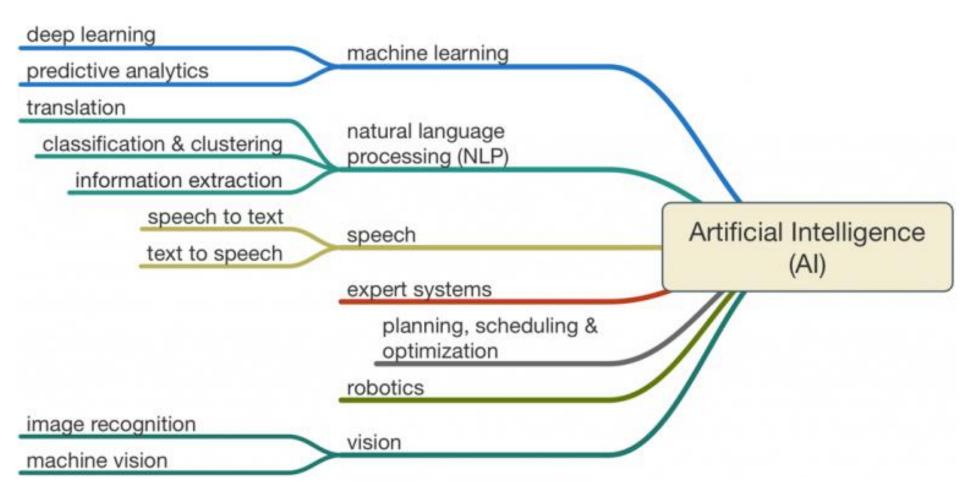
Um tipo de dado pouco utilizado pelas empresas são os dados de voz.

Quando estes dados de voz são utilizados para análise é comum que sejam transformados para o formato de texto, processo frequentemente denominado *speech-to-text*.



Macrovisão





Um possível caso de uso







TECNOLOGIA mais avançada do mercado hoje em dia...







... junto com uma METODOLOGIA orientada para os resultados

I. Entendimento da situação

II. Geração de Valor

- Definir objetivos de negócio
- 2. Criar e validar categorias
- 3. Análise de Categoria
- Sugestão causas raiz
- Dados diversos e priorização de chamadas
- Escutas focalizadas

- 4. Identificar Causas Raiz
 - 5. Análise e recomendações
 - Diretrizes de atuação
 - Impacto econômico
 - Implicações estratégicas, organizativas e operativas

Características do Modelo

> **CERTEZA NAS CATEGORIAS**

(mais de 85% de certeza)



CRUZAMENTO DE METADADOS



PRODUTIVIDADE NA ESCUTA



Um possível caso de uso



Aumentar RECEITAS



Aquisição de novos clientes

Retenção de clientes

Venda cruzada

Identificação de segmentos emergentes

Cobranças

Reduzir CUSTOS



Redução de chamadas / e-mails entrantes Melhoria de processos que geram dúvidas

Detecção de ocorrências de processos

Aumentar a QUALIDADE e a SATISFAÇÃO de clientes



Auditorias automáticas da qualidade de serviço

Análises de reclamações via oral e/ou escrita

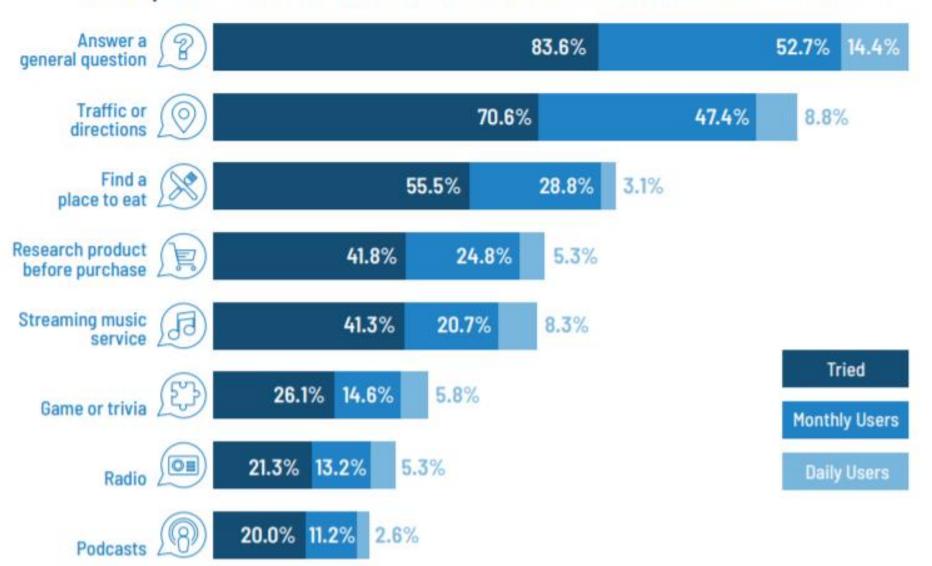
Análises de causas de fundo da insatisfação

Programas de formação personalizada de agentes Avaliação de fornecedores e equipas

Um mercado / negócio, jovem



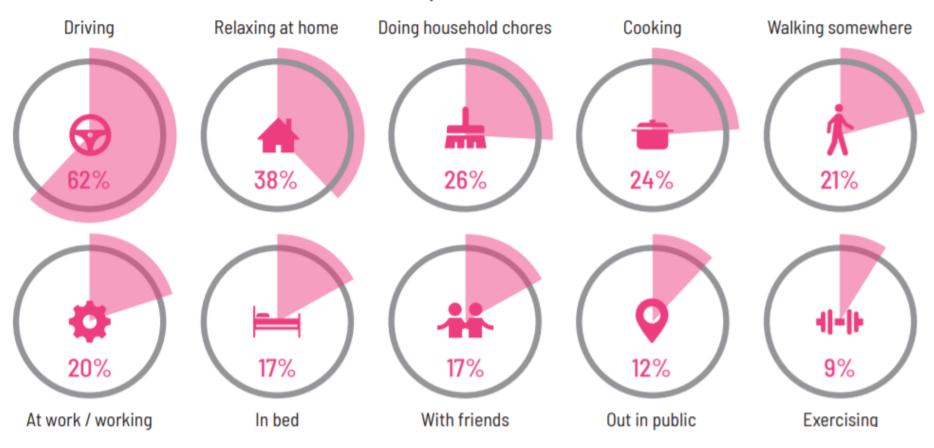
Smartphone Voice Assistant Information and Entertainment Use Cases



Um mercado / negócio, jovem



Common Use Cases for Smartphone Voice Assistant Users in the U.S.



Um mercado / negócio, jovem

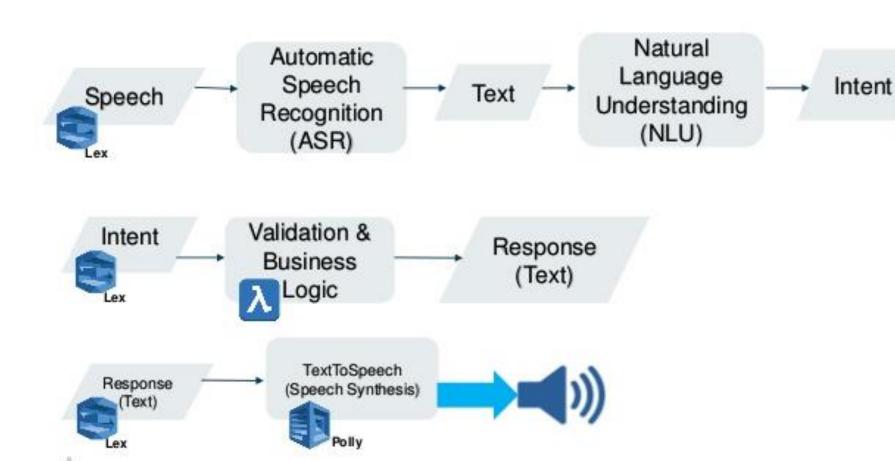


What Users Like About Voice Assistants on Smartphones

t is faster than touch and text t is faster than touch and text t is fun	t's hands free	
t is more convenient than touch and text 39% It is faster than touch and text 30% It is fun 19% Idon't like it 11% Idon't have an opinion on this 9% The answers are better than text-based search		60%
t is more convenient than touch and text	can do it while doing other things like driving or cooking	48%
t is faster than touch and text 30% t is fun 19% don't like it 11% don't have an opinion on this 9% The answers are better than text-based search		10 /0
t is fun 19% don't like it 11% don't have an opinion on this 9% The answers are better than text-based search	t is more convenient than touch and text	39 %
don't like it don't have an opinion on this 9% The answers are better than text-based search	t is faster than touch and text	30 %
don't have an opinion on this 9% The answers are better than text-based search	t is fun	19%
The answers are better than text-based search	don't like it	11%
	don't have an opinion on this	9%
	The answers are better than text-based search	6%

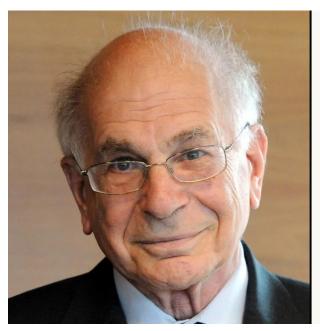
O modelo de voicebot

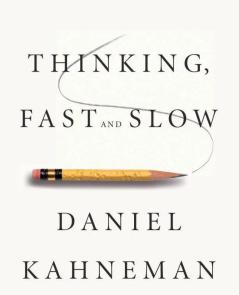




Duas formas de pensar







Pensamento vem do latim *pensare* e significa **pesar**, isto é **medir**, **avaliar** e **comparar**.

Sistema 1	Sistema 2
Inconsciente	Consciente
Não intencional	Intencional
Rápido	Lento
Sem esforço	Com esforço
Associativo	Lógico
Afetivo	Neutro
Rígido	Flexível
Intuitivo	Racional
Categórico	Individualizado

Duas formas de pensar



Existe 90% de chance de sucesso neste projeto

Existe 10% de chance de fracasso neste projeto



Robert Rosenthal





Na experiência, os alunos de uma escola americana foram submetidos a uma prova.

Rosenthal e sua equipe disseram aos 18 educadores do colégio que se tratava de um teste especial, desenvolvido na Universidade Harvard para analisar o potencial de desenvolvimento de cada criança.

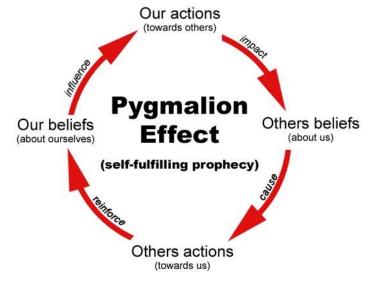
Mentira. Era apenas um reles teste de QI, sem nada de especial.



Robert Rosenthal







No final do ano escolar, a equipe de Rosenthal voltou à escola e repetiu o teste.

Os alunos que haviam sido falsamente diagnosticados como gênios haviam ganho, em média, 3,8 pontos de QI a mais que os demais.

O resultado foi ainda mais surpreendente entre alunos da primeira série: a diferença entre os ungidos e o resto foi de assombrosos 15,4 pontos de QI a mais.

A linguagem formando a realidade



How language shapes the way we think



https://www.ted.com/talks/lera_boroditsky_how_language_shapes_the_way_we_think

Steven Pinker Psicólogo graduado na Universidade McGill. PhD pela Universidade de Harvard.

Professor de Harvard.

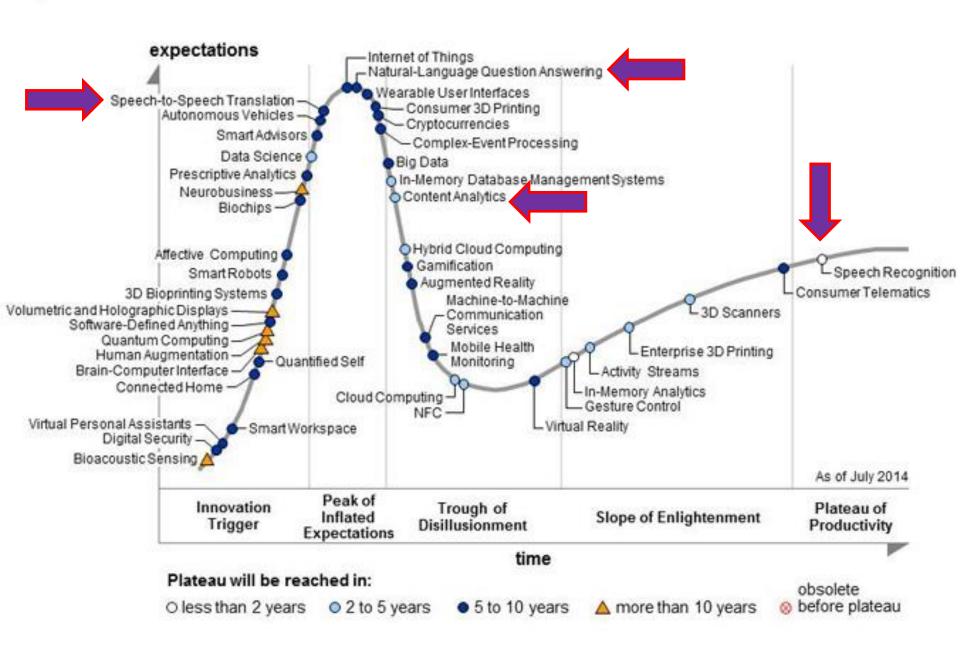
Lera Boroditsky Graduada e PhD por Stanford em Psicologia Cognitiva Professora da UCSD (University of California, San Diego)

What our language habits reveal



Para onde caminhamos

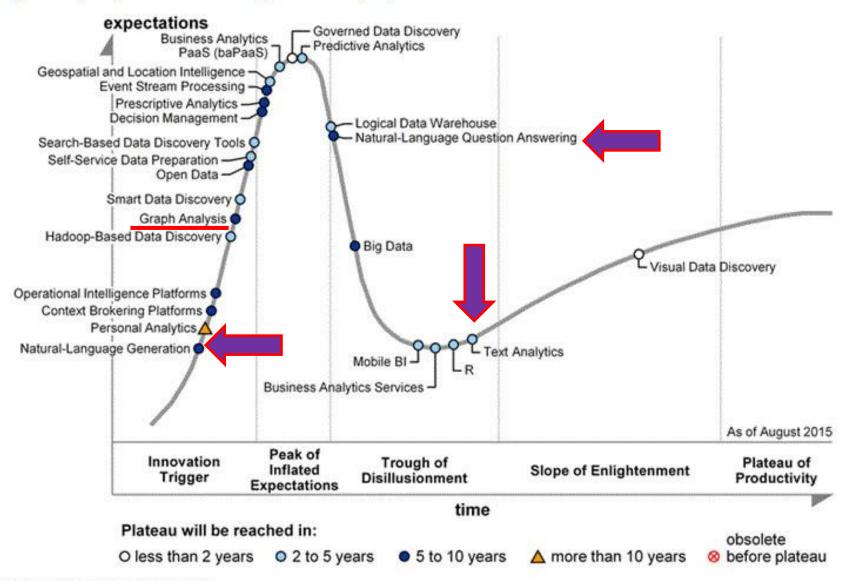




Uma visão do passado sobre o futuro



Figure 1. Hype Cycle for Business Intelligence and Analytics, 2015

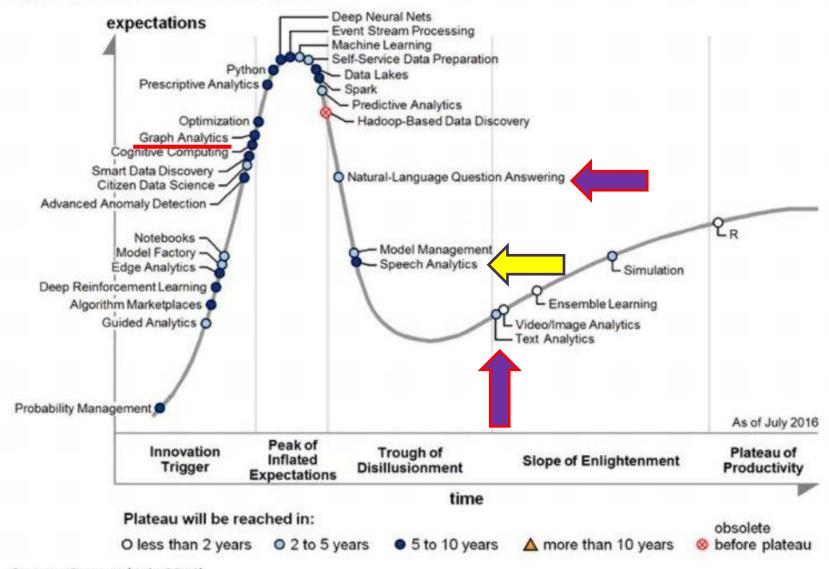


Source: Gartner (August 2015)

Uma visão do passado sobre o futuro



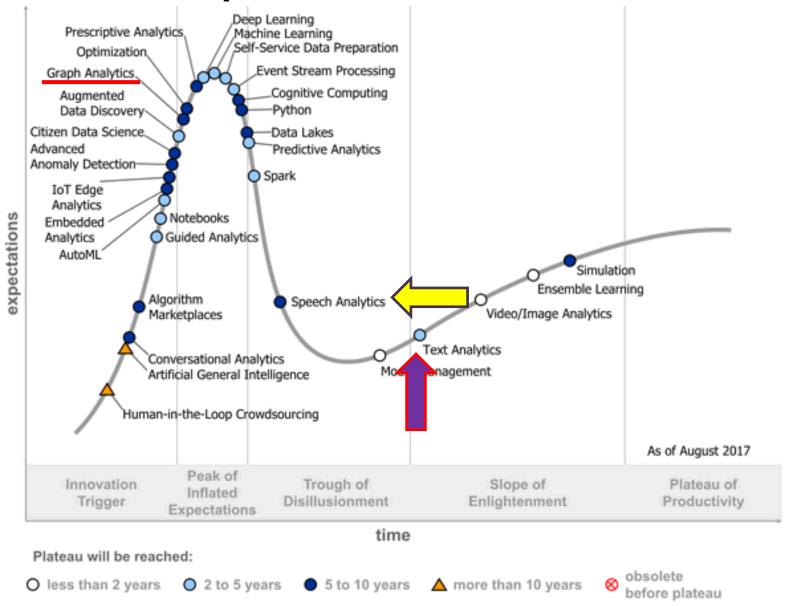
Figure 1. Hype Cycle for Data Science, 2016



Source: Gartner (July 2016)

Uma visão do passado sobre o futuro





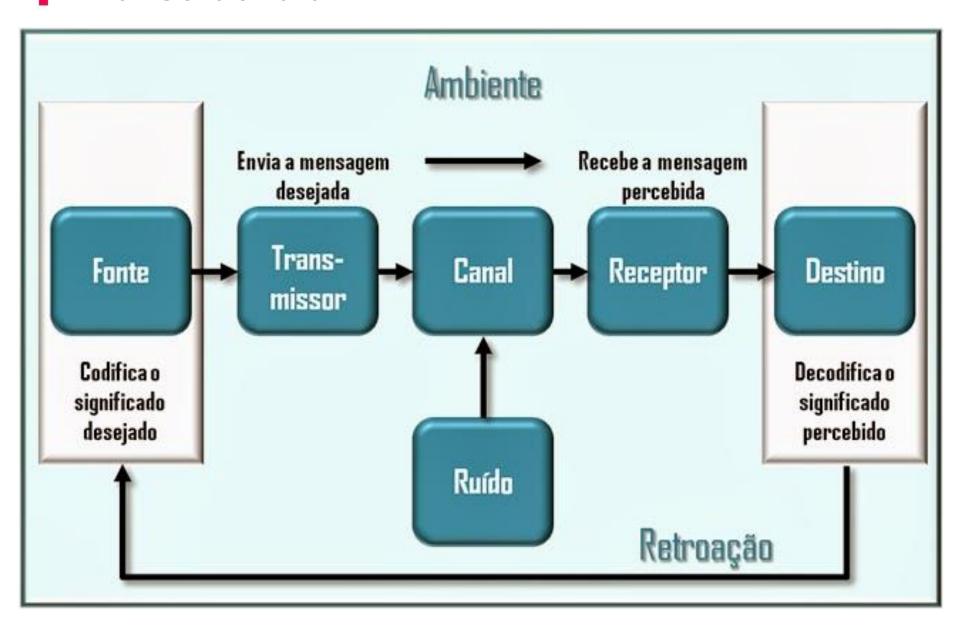
Artificial Intelligence & Machine Learning FIAP



Os dados

Análise de fala

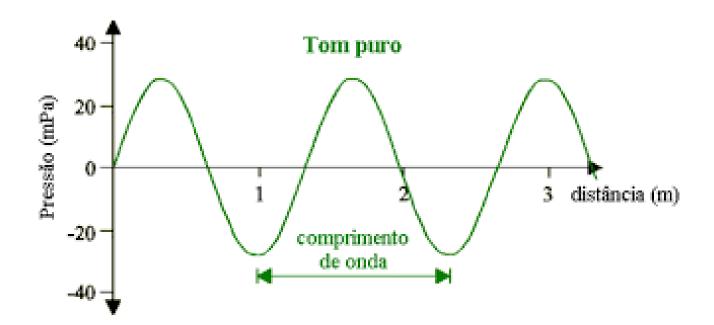




O som que ouvimos

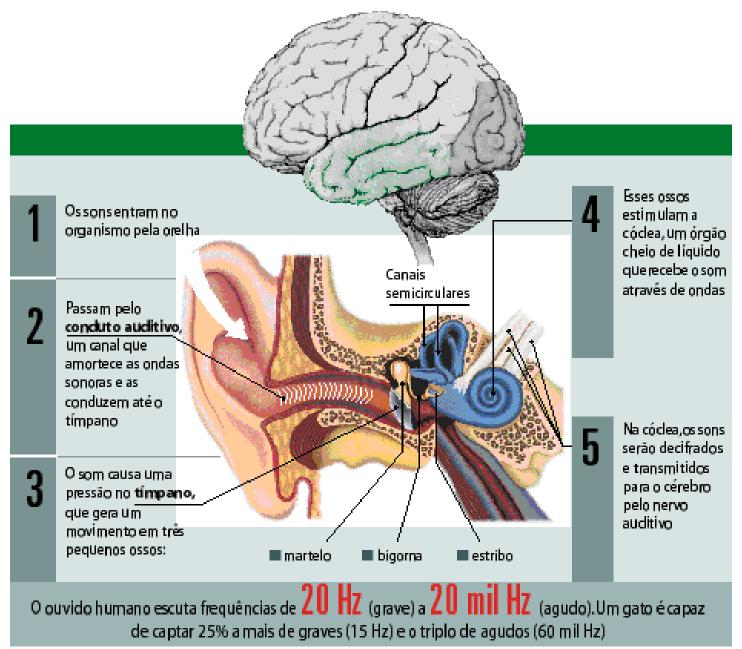


Todos os sons que ouvimos com nossos ouvidos **são ondas de pressão no ar**.



Como decodificamos o som





Onde no nosso cérebro é decodificado



Cérebro

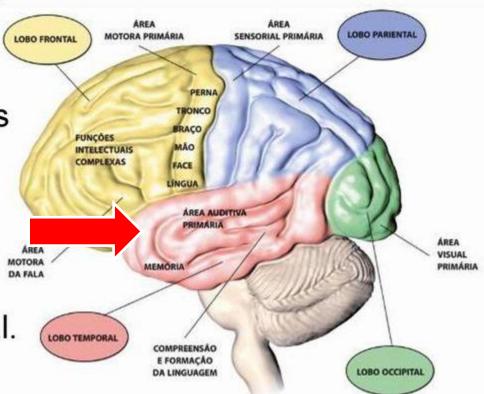
 O córtex cerebral compreende quatro grandes áreas:

Lobo Frontal: área
 responsável pela fala, pelas
 funções intelectuais e pela
 coordenação motora.

 Lobo Parietal: área sensorial e gustativa.

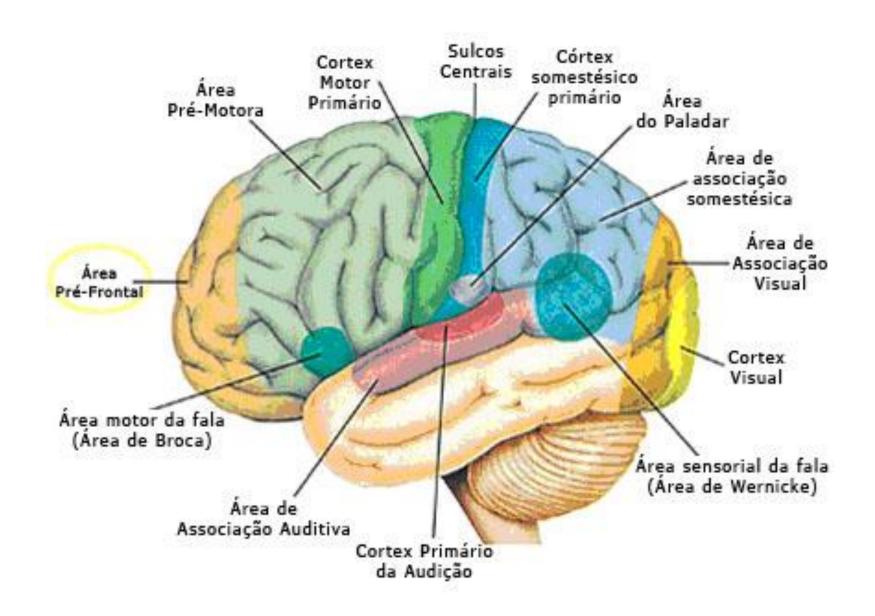
Lobo Occipital: área visual.

 Lobo Temporal: área auditiva e da memória.



Onde no nosso cérebro é decodificado





A difusão do som



Começando com a demonstração de Thomas Edison do primeiro fonógrafo em 1877, foi possível capturar essas ondas de pressão em um meio físico e depois reproduzi-las mais tarde regenerando as mesmas ondas de pressão.



Os dados são do século XIX

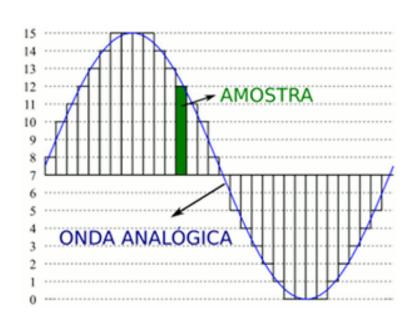


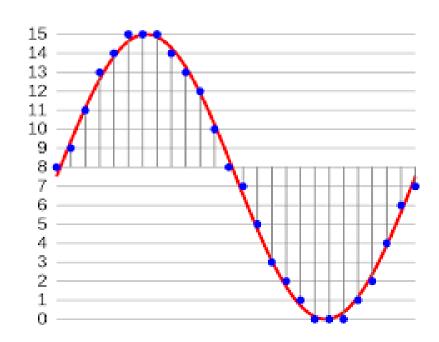
Sampling Rate	Fonte
8.000 hz	Telefone, walkie-talkie, microfone e comunicação wireless
22.000 hz	Radio
44.100 hz	Audio CD
48.000 hz	DVD, TV Digital
192.000 hz	Blue Ray, HD DVD (High Definition Digital)

Taxa de Amostragem



Taxa de amostragem é a quantidade de amostras de um sinal analógico coletadas em uma determinada unidade de tempo, para conversão em um sinal digital. Sendo uma frequência, é comumente medida em Hertz (Hz).

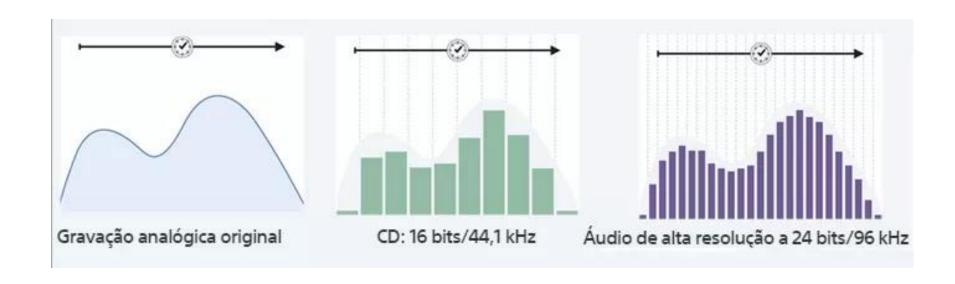




Taxa de Amostragem

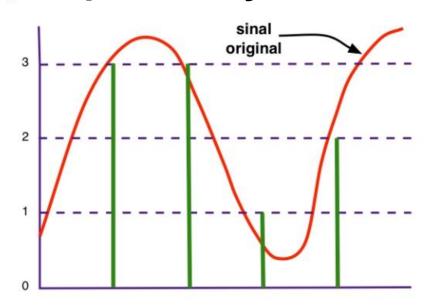


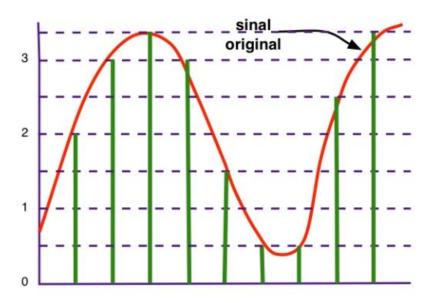
Quanto maior for a taxa de amostragem, mais medidas do sinal serão realizadas em um mesmo intervalo de tempo, e assim, maior será a fidelidade do sinal digital em relação ao sinal analógico.

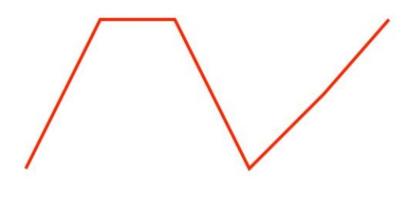


Representação do som







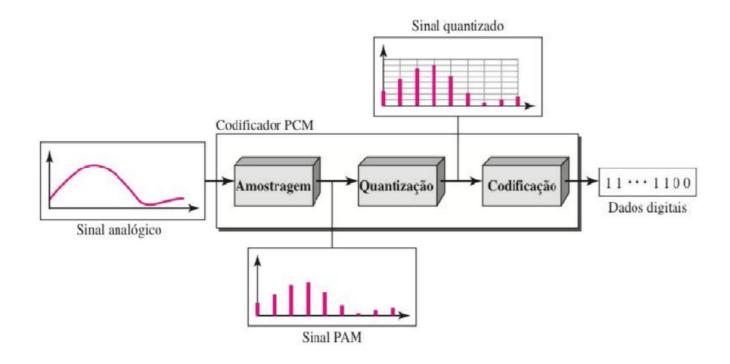


sinal digitalizado





O som é representado por uma sequência de amostras que são representadas numericamente em código binário, numa sequência de bits.





A Resolução refere-se ao número de bits usados para representar cada amostra. Uma amostra representada por um único bit admite apenas dois valores: "0" ou "1".

1 bit	2 valores	6 dB
2 bits	4 valores	12 dB
3 bits	8 valores	18 dB
4 bits	16 valores	24 dB
5 bits	32 valores	30 dB
6 bits	64 valores	36 dB
12 bits	4.096 valores	72 dB
16 bits	65.536 valores	96 dB
24 bits	16.777.216 valores	144 dB



Se considerarmos uma representação com 3 bits ela poderá obter 8 valores diferentes (2^3 = 8): 000, 001, 010, 100, 110, 101, 011, 111. Um CD tem, por exemplo, uma resolução de 16 bits o que permite uma resolução binária com 65.534 (216) valores.

Quanto maior for o número de níveis por bit, maior será a sua qualidade de reprodução e de semelhança com o som analógico.

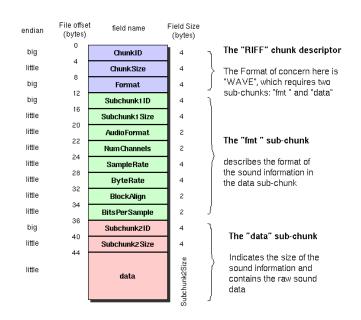


Fonte	Taxa Amostragem (em K Hz)	Bits por amostra
Telefone	8	8 kb / s
Radio AM	8	11 Kb / s
Radio FM	16	88.2 Kb / s
CD	16	176.4 Kb / s
DAT	16	192.0 Kb / s
Áudio DVD	24	1152.0 Kb / s

O formato para o mercado



WAV (Waveform Audio File Format), é um dos formatos mais conhecidos, não só pela sua grande utilização, mas também pelo seu peso e qualidade. Criado pela IBM e Microsoft.



O formato para o mercado



O formato WAV pode ser usado para qualidade máxima de áudio, podendo também ser editado e manipulado com relativa facilidade usando softwares. É, também, o mais recomendado para o trabalho de produção de áudio profissional dada a sua qualidade muito próxima do original. Por ser um formato sem compressão, o WAV ocupa um espaço muito grande de armazenamento.

Coder-decoder



Um codec é um dispositivo de <u>hardware ou programa de</u> <u>computador</u> capaz de <u>codificar ou descodificar um fluxo</u> <u>de dados de áudio</u> de acordo com um determinado tipo de arquivo de áudio ou áudio streaming.

O termo codec é uma combinação de coder-decoder (compressor/descompressor).



Coder-decoder



O **PCM** é um codec capaz de trabalhar com altas frequências de amostragem e resolução em bits. Ele é ideal quando o assunto é qualidade do áudio, trabalhando sem que o sinal seja perdido.

Existem outros codecs como o ADPCM ou o GSM-FR usado no CCO. Cada codec possui um range e qualidades dado seus propósitos.

O PCM pode ainda compactado utilizando dois métodos conhecidos como A-Law e U-Law.

Coder-decoder



Podem ser "Sem perdas" ou "com perdas"

Sem perdas (lossless) – codificam imagem ou som para comprimir o arquivo sem alterar a qualidade original. Se o arquivo for descomprimido manterá a mesma qualidade do original

Com perdas (lossy) – codificam imagem ou som, gerando uma certa perda de qualidade com a finalidade de alcançar maiores taxas de compressão.

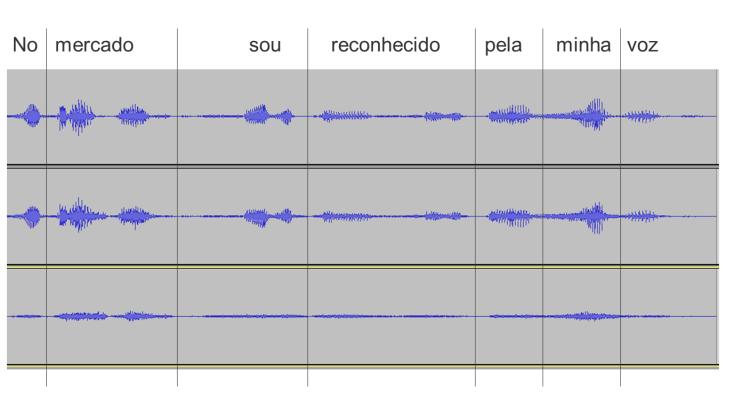


A compressão de áudio digital traz aos dias de hoje a possibilidade possuir arquivos sonoros de boa qualidade mas que não ocupem muito espaço.

A evolução da tecnologia permitiu a eliminação de sons que o ouvido humano não reconhece, levando assim a redução do espaço.

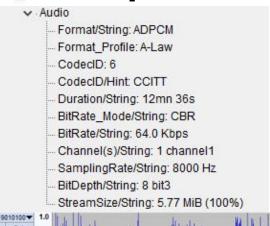
Compressão destrutiva - Exemplo



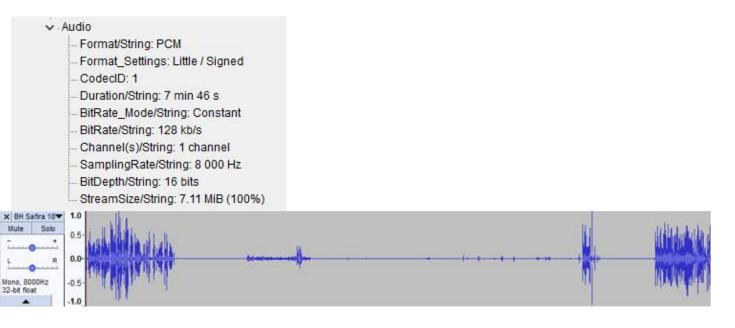


- (1) G711 8 KHz 8 Bits Sem compressão – Linha telefônica Brasil
- (2) Mesmo áudio anterior com compressão MP3 16KBits/s
- (3) = (1) (2) Diferença entre a forma de onda original e a comprimida = Perda de informações



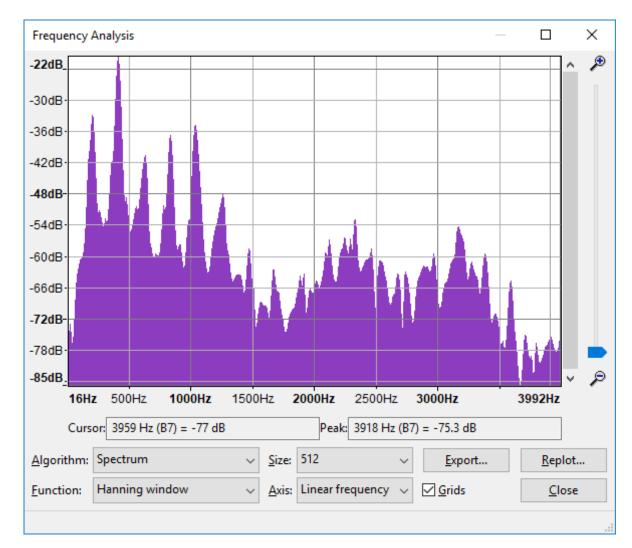






Exemplo 1 – Fonema "no" 8 Khz 8bits

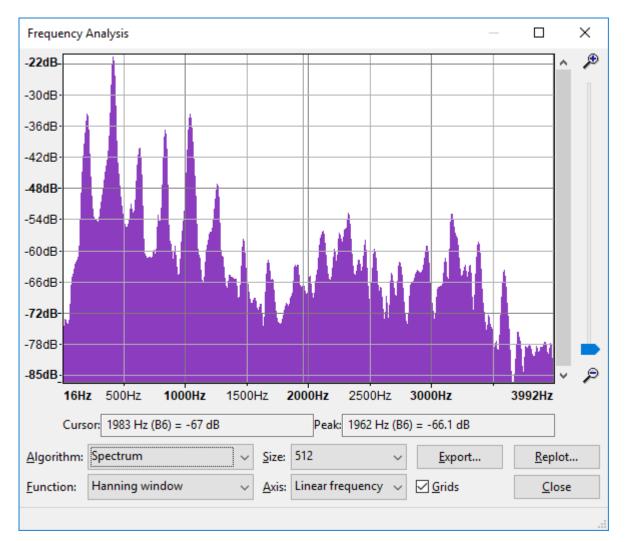




Análise de espectro do fonema "no" A maior frequência encontrada na pronúncia do fonema "no" gravada sem compressão

Exemplo 2 – Fonema "no" Mp3 16 Kbits

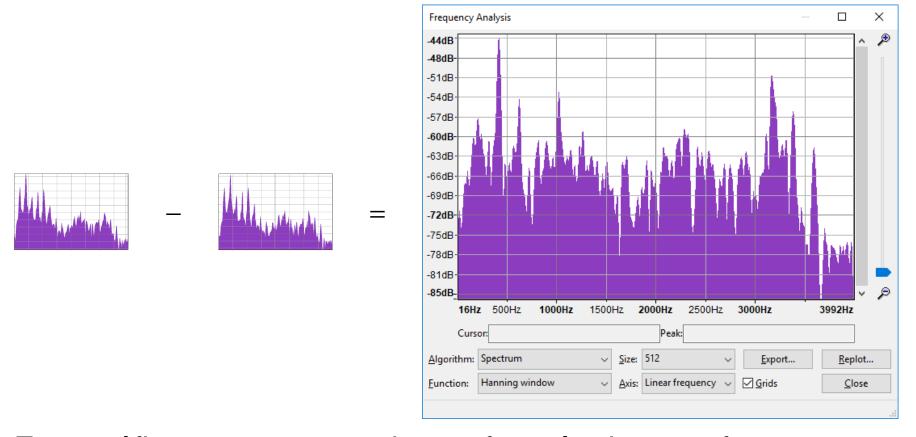




Nítida diferença na forma de onda, o fonema no perdeu algumas informações consideradas supérfluas para compreensão pelo ouvido humano.

Exemplo 3 – Fonema "no" Original subtraído do comprimido MP3





Este gráfico representa todas as frequências que foram atenuadas pela compressão MP3 e seu devido grau de atenuação.

TotalTime : 00:02:19.2030000



```
Audio FileName: 20190105023436-1101-082-157401-11978005000-00-21018@172.29.4.15060.wav
Encoding : ALaw
Channels: 1
SampleRate: 8000
BitsPerSample : 8
Length: 1108896
Length Formated: 1.06 MB
TotalTime: 00:02:18.6120000
Audio FileName: C0002-20190101203313.wav
Encoding: Gsm610
Channels: 1
SampleRate: 8000
BitsPerSample : 0
Length: 223466
Length Formated: 218 KB
TotalTime: 00:02:17.5180000
Audio FileName: wav03.wav
Encoding: Pcm
Channels: 1
SampleRate: 8000
BitsPerSample : 16
Length: 2227254
Length Formated: 2.12 MB
```



```
Audio FileName: 20190105022740-1101-104-157397-21998436042-00-21020@172.29.4.15060.wav
Encoding : ALaw
Channels: 1
SampleRate: 8000
BitsPerSample : 8
Length: 1386816
Length Formated: 1.32 MB
TotalTime : 00:02:53.3520000
Audio FileName: C0017-20190101070507.wav
Encoding : Gsm610
Channels: 1
SampleRate: 8000
BitsPerSample : 0
Length : 277416
Length Formated: 271 KB
TotalTime : 00:02:50.7180000
Audio FileName: wav02.wav
Encoding: Pcm
Channels: 1
SampleRate: 8000
BitsPerSample : 16
Length: 2796818
Length Formated: 2.67 MB
```

TotalTime : 00:02:54.8010000



```
Audio FileName: 20190105025047-1104-016-141428-31993389116-00-21021@172.29.4.15060.wav
Encoding : ALaw
Channels: 1
SampleRate: 8000
BitsPerSample: 8
Length: 2432640
Length Formated: 2.32 MB
TotalTime: 00:05:04.0800000
Audio FileName: C0001-20190101160109.wav
Encoding: Gsm610
Channels: 1
SampleRate: 8000
BitsPerSample: 0
Length: 492826
Length Formated: 481 KB
TotalTime : 00:05:03.2780000
Audio FileName: wav01.wav
Encoding: Pcm
Channels: 1
```

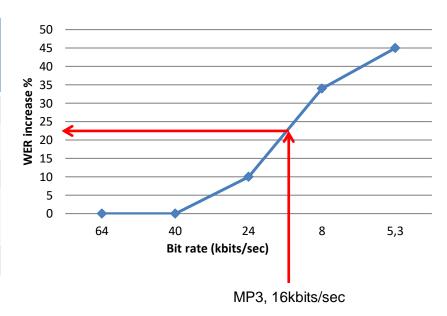
SampleRate: 8000 BitsPerSample : 16 Length: 4858304

Length Formated: 4.63 MB TotalTime : 00:05:03.6440000

Impactos da compressão na acuracidade FIAP de transcrição

Compressão vs G.711 – Sem compressão

Codec	Bit rate (kb/s)	Relative loss (WER increase) compared to G.711 (mismatch)
G.726	40	0%
G.726	32	1%
G.726	24	10%
G.729 (CS-ACELP)	8	34%
G.723	5.3	45%



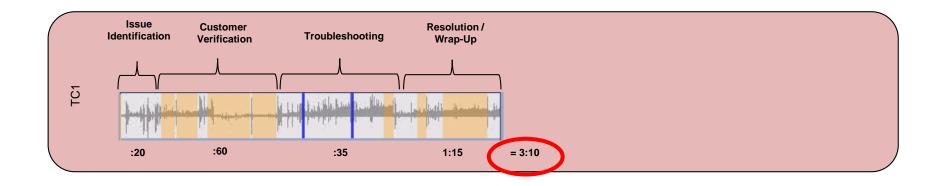
Treinamento acústico da solução pode mitigar o impacto da compressão

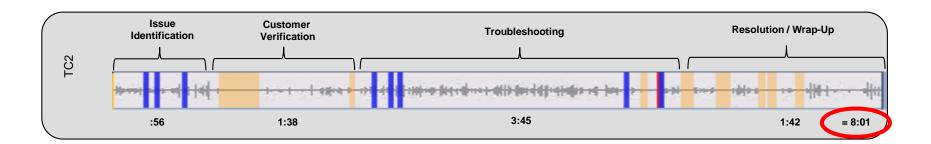
Impactos da compressão na acuracidade FIAP de transcrição

Codec	Data Rate (Kbps)	Representative Voice Quality (MOS)	Delay (ms)	Complexity (MIPS)
G.711 PCM	64.0	4.3	0.125	0
G.721 ADPCM	32.0	4.1	0.125	6.5
G.726 Multirate ADPCM	16 - 40	2.0 - 4.3	0.125	6.5
G.723 MP-MLQ ACELP	5.3, 6.3	4.1	70	25
G.728 LD-CELP	16.0	4.1	2	37.5
G.729 CS-ACELP	8.0	4.1	20	34
G.729a CS-ACELP	8.0	3.4	20	17

Modelo genérico. Funciona?







Áudios



As gravações devem ser fornecidas conforme orientação a seguir:

- Sem qualquer efeito de compressão ou encriptação
- Channel estéreo
- Formato wav
- Sample Rate: 8kHz
- Precision: 16bits
- Bit Rate: 128k
- O importante é o Sample Encoding tem que ser "16-bit Signed Integer PCM"

```
Channels : 1
Sample Rate : 8000
Precision : 16-bit
Duration : 00:00:08.96 = 71680 samples ~ 672 CDDA sectors
File Size : 143k
Bit Rate : 128k
Sample Encoding: 16-bit Signed Integer PCM
```

Metadados dos Áudios



O metadado deve possuir no mínimo os seguintes dados sobre cada ligação:

- Id único para correlacionar o metadado com a gravação
- Data (dd/mm/yyyy)
- Hora (hh:mm:ss)
- Ramal gravado
- Campanha
- Login Operador,
- Site,
- Gravadora (discador)
- DDD do originador e se possível o número de A inteiro
- Duração da chamada (millisegundos)
- Tamanho da gravação (kb)
- Outcome (detalhado nos dois slides seguintes)
- Para os casos de transferências de ramais permitir a rastreabilidade da ligação entre os ramais e gravações

Ainda sobre dados...





MediaInfo é um programa grátis que permite ter acesso a informações sobre os arquivos de áudio. Muito útil para avaliar o arquivo.



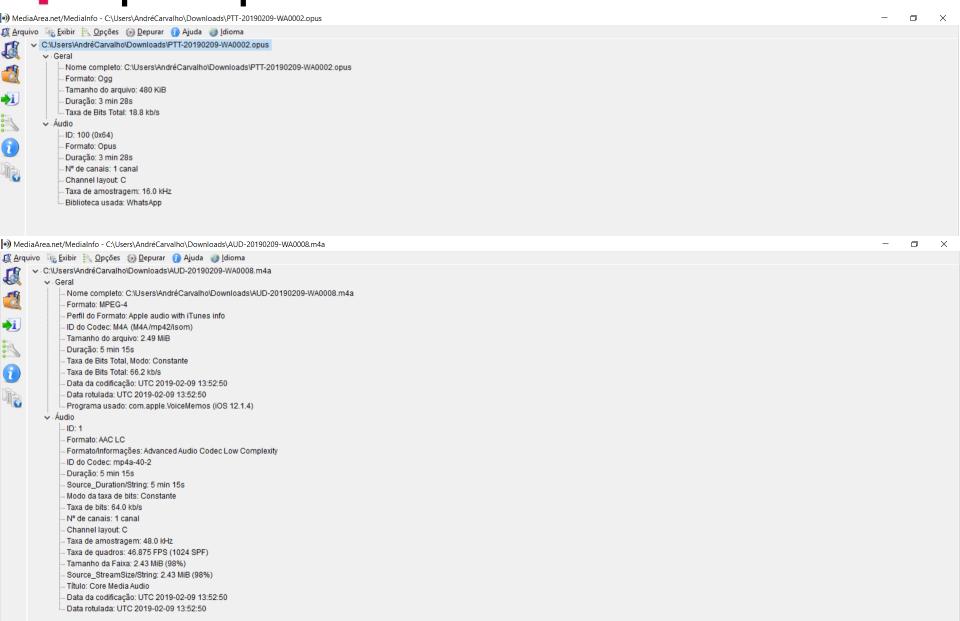






Arquivos para o trabalho final



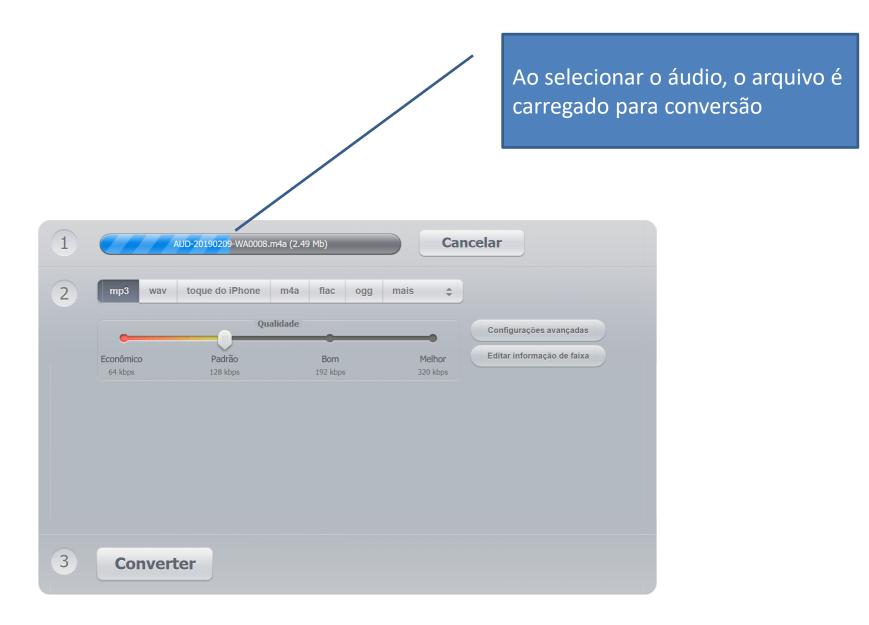




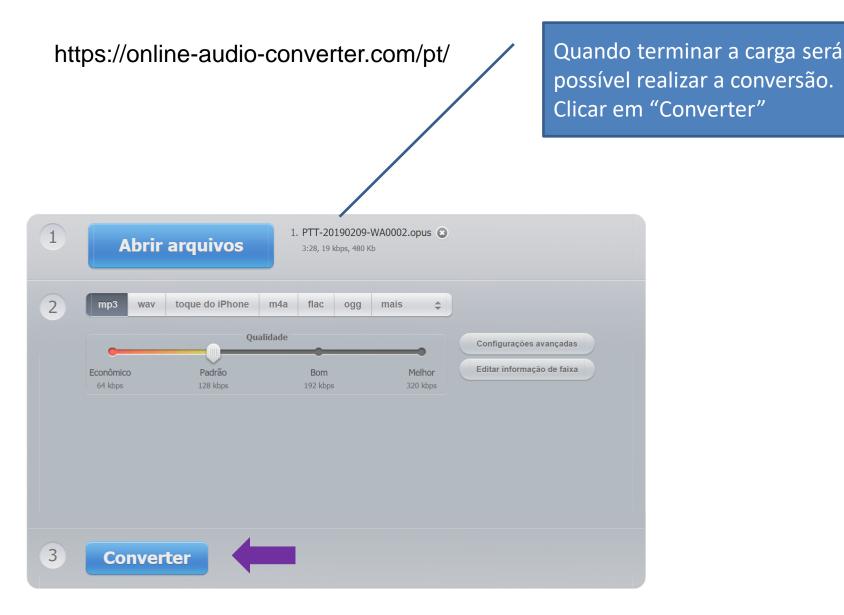
https://online-audio-converter.com/pt/



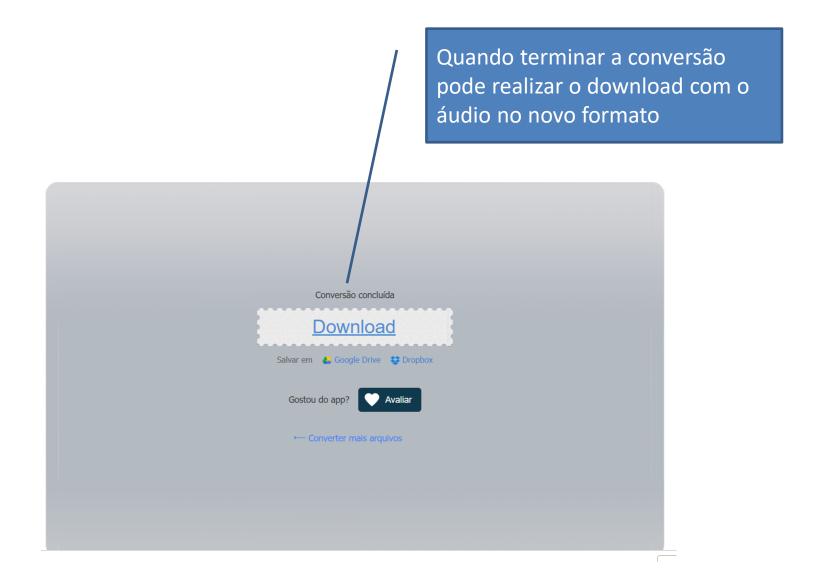






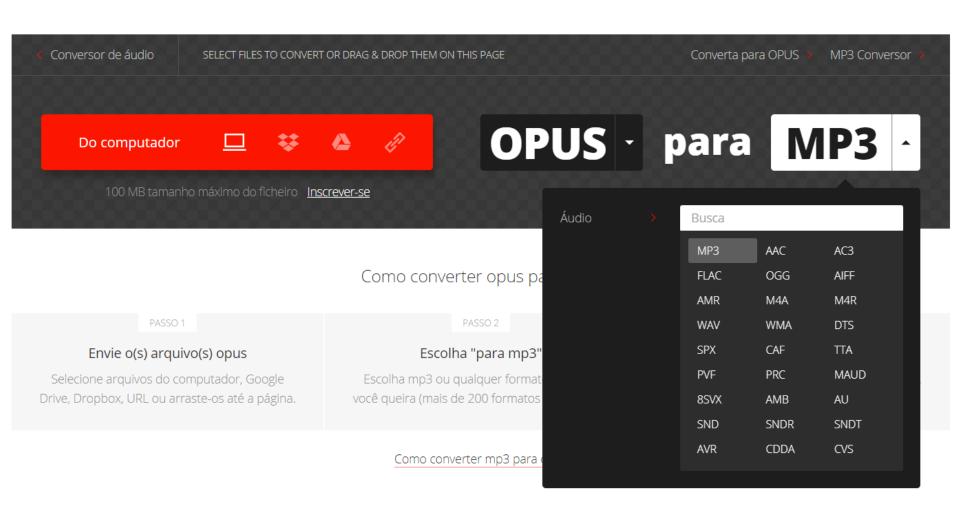








https://convertio.co/pt/opus-mp3/



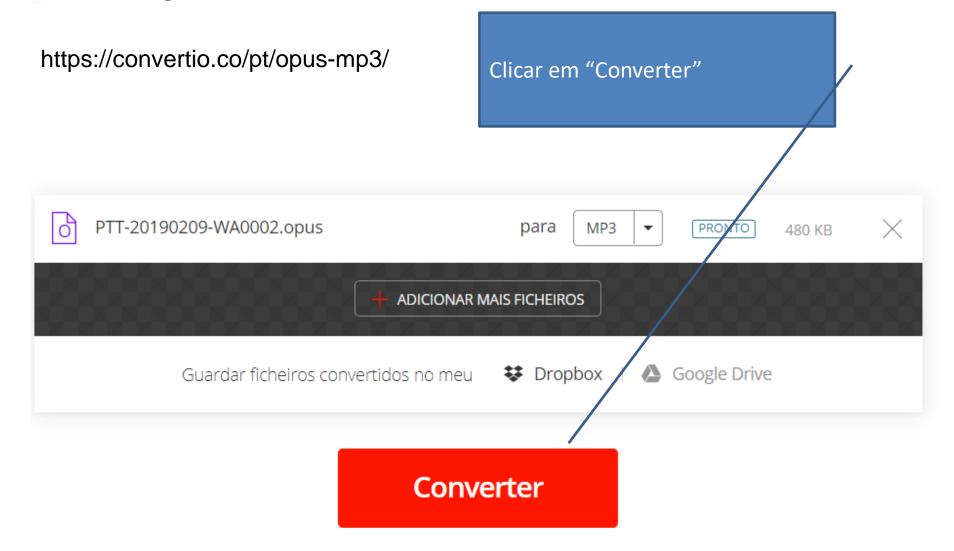




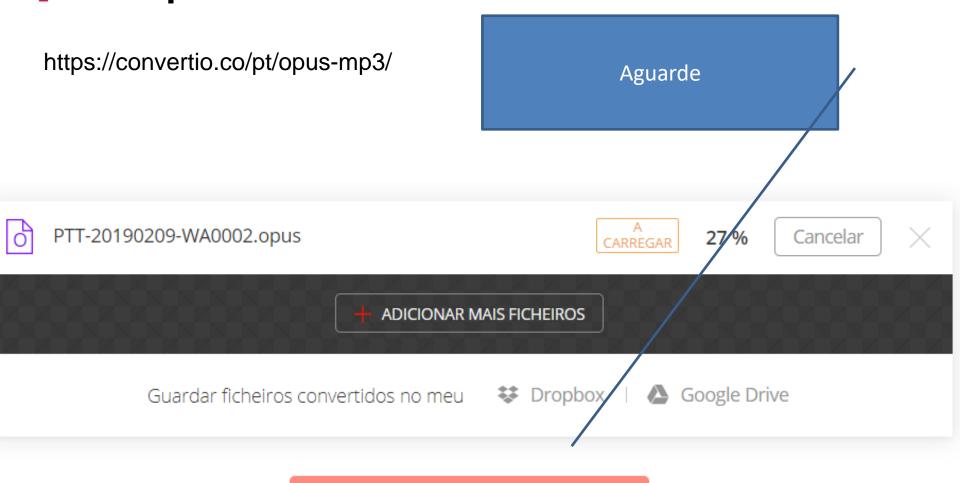


https://convertio.co/pt/opus-mp3/ Selecione o formato desejado Converta para OPUS Conversor de áudio SELECT FILES TO CONVERT OR DRAG & DROP THEM ON THIS PAGE MP3 Conversor OPUS para Do computador 100 MB tamanho máximo do ficheiro Inscrever-se Áudio MP3 AAC AC3 Como converter opus pa FLAC OGG AIFF **AMR** M4A M4R WAV WMA DTS SPX CAF TTA Envie o(s) arquivo(s) opus Escolha "para mp3" **PVF** PRC MAUD Selecione arquivos do computador, Google Escolha mp3 ou qualquer format Drive, Dropbox, URL ou arraste-os até a página. você queira (mais de 200 formatos 8SVX AMB ΑU SND **SNDR** SNDT CDDA CVS AVR Como converter mp3 para









Por favor, Aguarde





Converter





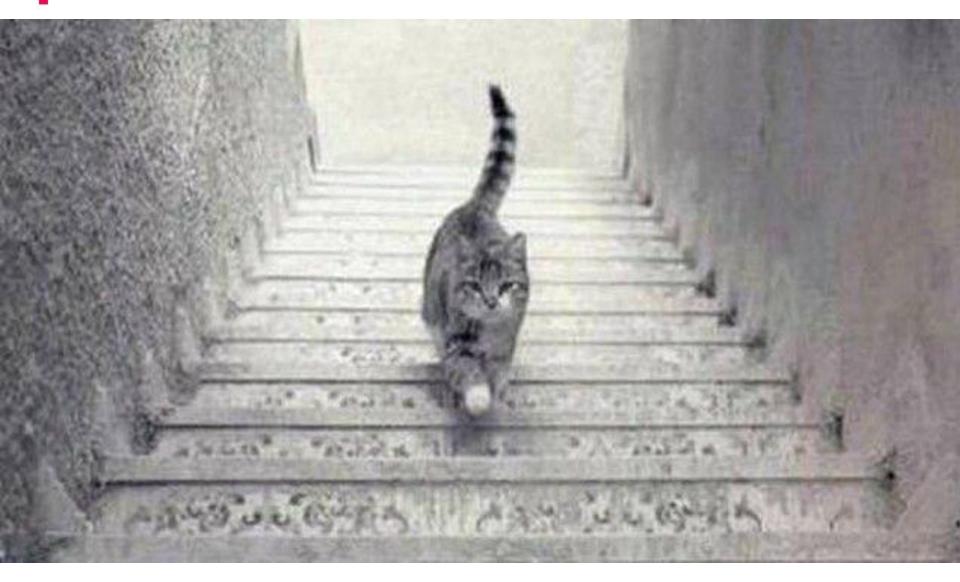
Converter



Obrigado por enquanto!

Descendo ou subindo?









Copyright © 2016 Prof.

Todos direitos reservados. Reprodução ou divulgação total ou parcial deste documento é expressamente proibido sem o consentimento formal, por escrito, do Professor (autor).