ANALISADOR VOCAL AI

Sistema de Análise de Características Vocais para Replicação Técnica

Extração automatizada de features vocais

Análise baseada em processamento de sinais

Geração de relatórios técnicos e práticos

Versão 5.0 | Documentação Técnica | Janeiro 2025

Google Colab Implementation

SALVAR COMO PDF

Índice

6. Limitações e Considerações

7. Referências Técnicas

1. Visão Geral do Sistema

1.1 O Que o Sistema Faz

> **Em termos simples:** O sistema pega um arquivo de áudio (uma música cantada) e mede diversas características da voz, gerando um relatório que explica tecnicamente como aquela pessoa está cantando.
> **Objetivo:** Fornecer dados objetivos e quantificáveis sobre técnica vocal, permitindo análise comparativa e replicação de estilos.

1.2 Componentes Principais

Componente	Função	Tecnologia
Separação de Fonte	Isola vocais de instrumentos	Demucs (deep learning)
Transcrição	Converte áudio em texto com timestamps	OpenAI Whisper
Análise Acústica	Extrai features vocais	Librosa, Praat/Parselmouth
Geração de Relatório	Traduz dados técnicos para linguagem prática	Python (processamento de texto)

1.3 Entrada e Saída

> **INPUT:**
>
> - **Formato de áudio:** WAV (apenas)
> - **Tipo de conteúdo:** Vocais isolados (sem instrumentos) ou música completa
> - **Qualidade recomendada:** 44.1 kHz, 16-bit ou superior
> - **Duração:** Ilimitada (testado com músicas de até 40 minutos)

- *Nota: Tempo de processamento depende dos recursos do Google Colab disponíveis no momento*

OUTPUT:

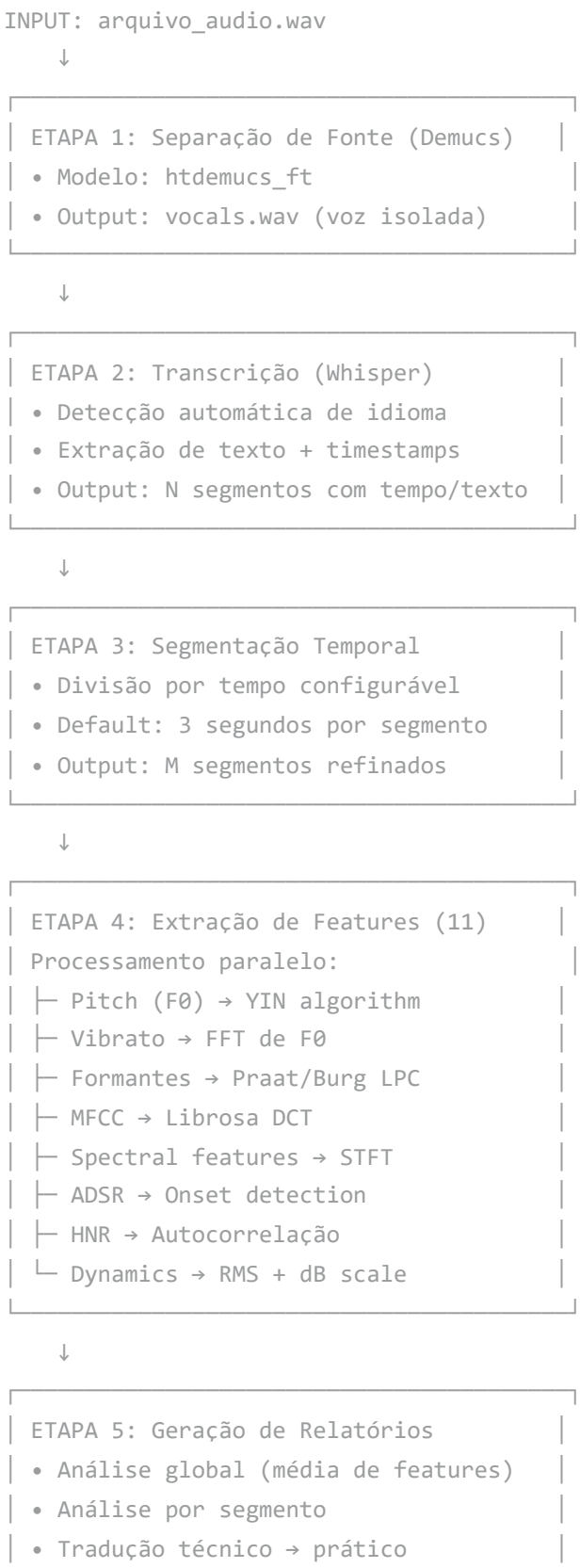
- Arquivo TXT com relatório em duas partes:
  - **Parte 1:** Guia prático (linguagem natural)
  - **Parte 2:** Dados técnicos por segmento

1.4 Casos de Uso

- **Cantores:** Análise de referências para estudo técnico
- **Professores de Canto:** Diagnóstico objetivo de alunos
- **Produtores:** Direcionamento técnico em sessões de gravação
- **Pesquisadores:** Estudos comparativos de estilos vocais
- **Análise de Progresso:** Comparação antes/depois de treinamento

2. Arquitetura e Pipeline

2.1 Fluxo de Processamento



```
└──────────────────────────────────────────┘
    ↓
OUTPUT: relatorio.txt
```

2.2 Dependências do Sistema

Biblioteca	Versão	Uso Principal
openai-whisper	latest	Transcrição e detecção de idioma
demucs	4.0+	Separação vocal/instrumental
librosa	0.10+	Análise de áudio (F0, MFCC, espectral)
praat-parselmouth	0.4+	Formantes, HNR, jitter, shimmer
pytorch	2.0+	Backend para modelos de ML
scipy	1.10+	Processamento de sinais
numpy	1.24+	Computação numérica

> 💡 **Sobre Tempo de Processamento:**
>
> O tempo de processamento varia conforme os recursos disponíveis no Google Colab no momento da execução:
>
> - **Com GPU alocada:** Processamento mais rápido (~5-10 min para música de 3-4 min)
> - **Apenas CPU:** Processamento mais lento (~15-30 min para música de 3-4 min)
> - **Músicas longas (20-40 min):** Processamento proporcional à duração
>
> *Nota: Google Colab aloca recursos dinamicamente. GPU não está sempre disponível na versão gratuita.*

3. Features Analisados

1 Pitch (F0) - Frequência Fundamental

3.1.1 Definição Técnica

> **Feature:** Fundamental Frequency (F0)
> **Algoritmo:** YIN (autocorrelação com threshold adaptativo)
> **Implementação:** `librosa.pyin()`
> **Range de detecção:** 80 Hz (C2) a 1200 Hz (D6)
> **Taxa de amostragem:** Hop length = 512 samples
> **Output:** Array de frequências (Hz) + flags de vocalização
> **Precisão:** ±0.5 Hz

3.1.2 O Que Significa (Linguagem Natural)

> **Analogia:** É como medir qual tecla do piano está sendo tocada. Se alguém canta uma nota "Lá", o F0 será aproximadamente 440 Hz.
> **Por que importa:** Determina se a voz é grave (baixo Hz) ou aguda (alto Hz). É o parâmetro mais básico e essencial - cantar na altura errada torna impossível replicar qualquer performance.
> **Exemplo prático:**
> - Voz masculina grave: ~110 Hz (A2)
> - Voz masculina média: ~130 Hz (C3)
> - Voz feminina média: ~220 Hz (A3)
> - Voz feminina aguda: ~440 Hz (A4)

3.1.3 Dados Extraídos

> **Estatísticas calculadas:**
> - `f0_media` : Média aritmética de F0 (Hz)
> - `f0_mediana` : Valor central (Hz)
> - `f0_std` : Desvio padrão (variabilidade)

- `f0_min` : Frequência mais grave (Hz)
- `f0_max` : Frequência mais aguda (Hz)
- `f0_range` : Extensão total (max - min)

3.1.4 Interpretação

F0 Médio (Hz)	Classificação	Registro Típico
< 150	Grave (Baixo/Contralto)	Voz de peito predominante
150 - 250	Médio (Barítono/Mezzo)	Voz mista frequente
250 - 400	Agudo (Tenor/Soprano)	Voz de cabeça/mista
> 400	Muito Agudo	Falsete/head voice

2 Vibrato - Modulação Periódica de F0

3.2.1 Definição Técnica

> **Feature:** Vibrato (modulação periódica da frequência fundamental)
> **Algoritmo:** FFT do contorno de F0
> **Implementação:** `np.fft.fft(f0 - mean(f0))`
> **Parâmetros medidos:**
> • **Taxa (Hz):** Frequência da oscilação (oscilações/segundo)
> • **Extensão (cents):** Amplitude da modulação (1 semitom = 100 cents)
> • **Regularidade:** Coeficiente de variação temporal
> **Threshold de detecção:** Taxa > 3 Hz E Extensão > 20 cents
> **Range típico:** 3-8 Hz (taxa), 20-200 cents (extensão)

3.2.2 O Que Significa (Linguagem Natural)

> **Analogia:** É a "tremidinha" que cantores fazem em notas longas. Como quando você segura uma nota e ela oscila levemente para cima e para baixo.
> **Por que importa:** O vibrato define o estilo de canto. Ópera usa vibrato lento e amplo. Pop moderno frequentemente não usa vibrato. Gospel usa vibrato rápido.
> **Visualização:**
> - Sem vibrato: _____ (linha reta)
> - Vibrato lento (3.5 Hz): ∿∿∿∿ (ondas largas)
> - Vibrato rápido (7 Hz): ≈≈≈≈≈ (ondas estreitas)

3.2.3 Dados Extraídos

> **Output do sistema:**
> - `presente` : Boolean (vibrato detectado ou não)
> - `taxa_hz` : Frequência de oscilação (Hz)
> - `extensao_cents` : Amplitude da modulação (cents)
> - `regularidade` : 1/(1 + std(diff(f0))) → [0,1]

3.2.4 Interpretação

Taxa (Hz)	Extensão (cents)	Característica	Estilos Típicos
3.0 - 4.5	50 - 100	Lento e amplo	Ópera, Música Clássica
4.5 - 6.0	40 - 80	Moderado	Jazz, Soul, Baladas
6.0 - 8.0	30 - 60	Rápido e tenso	Gospel, R&B
N/A	< 20	Ausente/Sutil	Pop moderno, Indie

3 Formantes - Ressonâncias do Trato Vocal

3.3.1 Definição Técnica

Feature: Formantes (picos de ressonância no espectro vocal)
Algoritmo: Linear Predictive Coding (LPC) via Algoritmo de Burg
Implementação: Praat via `parselmouth.Sound.to_formant_burg()`
Parâmetros:
- Número de formantes: 5 (F1, F2, F3, F4, F5)
- Frequência máxima: 5500 Hz
- Window length: 0.025 s
- Pre-emphasis: 50 Hz

Output: Frequência média, std, min, max para cada formante
Precisão: ±25 Hz por formante

3.3.2 O Que Significa (Linguagem Natural)

Analogia: Imagine sua boca e garganta como um instrumento musical (tipo uma flauta). Dependendo da FORMA que você faz com esses espaços, certas frequências ficam mais altas (ressoam). Essas frequências que ressoam são os formantes.
Por que importa: Formantes são o "DNA" da sua voz. Duas pessoas podem cantar a MESMA NOTA (mesmo F0), mas soarem completamente diferentes por causa dos formantes. É o que faz sua voz ter seu timbre único.
O que cada formante controla:

- **F1 (300-900 Hz):** Abertura vertical da boca (quão aberta está a mandíbula)
- **F2 (800-2500 Hz):** Posição horizontal da língua (frente/trás)
- **F3-F5:** Características individuais do timbre

3.3.3 Dados Extraídos

Para cada formante (F1 a F5):

- `Fi_media` : Frequência média (Hz)
- `Fi_std` : Variabilidade ao longo do tempo
- `Fi_min` : Valor mínimo observado

- `Fi_max` : Valor máximo observado

3.3.4 Interpretação - F1 (Abertura da Boca)

F1 (Hz)	Abertura	Vogais Típicas	Instrução Prática
< 400	Muito fechada	"i", "u"	Mandíbula quase fechada
400 - 550	Fechada	"e", "o"	Mandíbula levemente aberta
550 - 700	Média	"ê", "ô"	Abertura natural, confortável
700 - 850	Aberta	"é", "ó", "a"	Mandíbula bem aberta
> 850	Muito aberta	"á"	Mandíbula baixa, boca bem aberta

3.3.5 Interpretação - F2 (Posição da Língua)

F2 (Hz)	Posição da Língua	Timbre Resultante
< 1000	Muito para trás	Som muito escuro, "aveludado"
1000 - 1400	Trás	Som escuro, "redondo"
1400 - 1800	Centro/Neutro	Som equilibrado, natural
1800 - 2200	Frente	Som claro, "brilhante"
> 2200	Muito para frente	Som muito brilhante, "metálico"

4 Breathiness - Componente Aperiódico

3.4.1 Definição Técnica

```
Feature: Breathiness (proporção de ruído turbulento vs. harmônicos)
Algoritmo: Análise espectral em faixas de frequência
Cálculo: Energia(f > 4kHz) / Energia(f < 1kHz)
Implementação:
 stft = np.abs(librosa.stft(audio))
freqs = librosa.fft_frequencies(sr=sr)
energia_alta = sum(stft[freqs > 4000])
energia_baixa = sum(stft[freqs < 1000])
ratio = energia_alta / energia_baixa
Range: 0.0 (sem ar) a 1.0+ (muito aéreo)
Correlação física: Grau de adução das cordas vocais
```

3.4.2 O Que Significa (Linguagem Natural)

Analogia: É quanto ar está "vazando" enquanto você canta. Como a diferença entre falar normalmente (pouco ar) vs. sussurrar (muito ar).
Por que importa: Breathiness define se a voz é "íntima" ou "poderosa". Billie Eilish tem alto breathiness (voz sussurrada). Whitney Houston tem baixo breathiness (voz limpa e potente).
Física: Quando as cordas vocais não fecham completamente, ar escapa criando ruído. Quanto mais ar escapa, maior o breathiness.

3.4.3 Dados Extraídos

Output:

- `breathiness_ratio` : Valor numérico [0, 1+]
- `nivel` : Classificação textual (baixo/médio/alto)

3.4.4 Interpretação

Ratio	Nível	Característica	Exemplos de Artistas
< 0.1	Baixo	Voz limpa, comprimida, potente	Whitney Houston, Adele, Freddie Mercury
0.1 - 0.3	Médio	Voz natural, equilibrada	Maioria dos cantores pop
> 0.3	Alto	Voz aérea, sussurrada, íntima	Billie Eilish, Norah Jones, Lana Del Rey

5 Brilho Espectral - Centróide

3.5.1 Definição Técnica

> **Feature:** Spectral Centroid (centro de massa do espectro)
> **Algoritmo:** Média ponderada das frequências por amplitude
> **Fórmula:** Centroid = Σ(f × magnitude(f)) / Σ(magnitude(f))
> **Implementação:** `librosa.feature.spectral_centroid(y=audio)`
> **Features relacionados:**
> • Spectral Rolloff: Frequência onde 85% da energia está abaixo
> • Spectral Bandwidth: Dispersão em torno do centróide
> **Range típico:** 500 Hz (muito escuro) a 4000+ Hz (muito brilhante)

3.5.2 O Que Significa (Linguagem Natural)

> **Analogia:** É se sua voz é "clara/brilhante" como uma luz LED, ou "escura/aveludada" como luz de vela.
> **Por que importa:** Define a "cor" da voz. Uma voz com alto centróide espectral soa "clara", "incisiva", "metálica". Com baixo centróide soa "quente", "aveludada", "escura".
> **Onde você sente a ressonância:**
> - **Alto brilho:** Máscara facial (nariz, maçãs do rosto, testa)
> - **Baixo brilho:** Peito, garganta

3.5.3 Dados Extraídos

> **Output:**
> - `centroide_media` : Média do centróide (Hz)
> - `centroide_std` : Variabilidade
> - `rolloff_media` : Média do rolloff (Hz)
> - `bandwidth_media` : Largura espectral média (Hz)
> - `nivel` : Classificação (escuro/médio/brilhante)

3.5.4 Interpretação

Centróide (Hz)	Classificação	Característica Sonora	Aplicação Típica
< 1000	Muito Escuro	Som "fechado", "aveludado", grave	Blues, Jazz vocal
1000 - 1500	Escuro	Som quente, redondo	Soul, R&B clássico
1500 - 2500	Equilibrado	Som natural, balanceado	Pop, Rock
2500 - 3500	Brilhante	Som claro, projetado	Musical Theatre, Pop moderno
> 3500	Muito Brilhante	Som "metálico", incisivo, penetrante	Gospel, Belting

6 Tensão Vocal - Energia em Faixas Críticas

3.6.1 Definição Técnica

```
Feature: Vocal Strain/Tension (energia na faixa de "tensão")
Algoritmo: Análise espectral em 2-4 kHz
Cálculo: Energia(2kHz-4kHz) / Energia_total
Implementação:
 stft = np.abs(librosa.stft(audio))
mask = (freqs >= 2000) & (freqs <= 4000)
ratio = sum(stft[mask]) / sum(stft)
Range: 0.0 (relaxado) a 0.5+ (muito tenso)
Correlação física: Grau de constrição faríngea e adução glótica
```

3.6.2 O Que Significa (Linguagem Natural)

Analogia: É quão "apertada" ou "solta" está sua garganta quando você canta. Como a diferença entre fazer força para abrir uma tampa (tenso) vs. bocejar (relaxado).
Por que importa: A tensão vocal define o caráter da voz. Rock e gospel frequentemente usam alta tensão (som "gritado", "apertado"). Jazz e música clássica usam baixa tensão (som "solto", "fácil").
IMPORTANTE: Tensão excessiva pode causar danos às cordas vocais. Este parâmetro deve ser usado com cuidado e supervisão.

3.6.3 Dados Extraídos

Output:

- `tensao_ratio` : Valor numérico [0, 0.5+]
- `nivel` : Classificação (baixa/média/alta)

3.6.4 Interpretação

Ratio	Nível	Característica	Aplicação
< 0.15	Baixa	Garganta relaxada, som "fácil"	Jazz, Clássico, Baladas suaves

Ratio	Nível	Característica	Aplicação
0.15 - 0.30	Média	Tensão controlada, natural	Pop, R&B
> 0.30	Alta	Garganta comprimida, som "apertado"	Rock, Metal, Gospel power

> ⚠️ **AVISO MÉDICO:** Tensão vocal excessiva (>0.40) mantida por longos períodos pode causar nódulos, pólipos ou outras lesões nas cordas vocais. Este sistema mede, mas não recomenda níveis extremos. Consulte um fonoaudiólogo antes de tentar replicar técnicas de alta tensão.

7 Ataque (ADSR) - Envelope Temporal

3.7.1 Definição Técnica

Feature: ADSR Envelope (Attack, Decay, Sustain, Release)
Algoritmo: Análise de onset strength
Implementação:

```
envelope = librosa.onset.onset_strength(y=audio)
peak_idx = np.argmax(envelope)
attack_time = peak_idx / sr * hop_length
```

Parâmetros medidos:

- Attack time: Tempo de 0% a 100% de amplitude (segundos)
- Sustain level: Nível médio após o ataque

Range típico: 0.01s (muito rápido) a 0.5s+ (muito lento)

3.7.2 O Que Significa (Linguagem Natural)

Analogia: É como você COMEÇA cada nota. Como bater numa bateria (ataque duro, instantâneo) vs. ligar um dimmer de luz (ataque suave, gradual).
Por que importa: O ataque define o impacto emocional. Ataque rápido transmite energia, assertividade. Ataque lento transmite suavidade, contemplação.
Exemplos:

- **Ataque duro:** "BAM!" - Pop energético, Rock
- **Ataque suave:** "mmmmm..." - Baladas, Soul

3.7.3 Dados Extraídos

Output:

- `attack_time` : Tempo de ataque (segundos)
- `sustain_level` : Nível de sustentação (amplitude)
- `envelope_std` : Variabilidade do envelope

3.7.4 Interpretação

Attack Time (s)	Classificação	Sensação	Estilos Típicos
< 0.05	Muito Rápido/Duro	Impacto imediato, energia	Rock, Pop up-tempo, R&B
0.05 - 0.15	Moderado	Natural, equilibrado	Pop, Country
> 0.15	Lento/Suave	Suavidade, gradual, contemplativo	Baladas, Soul, Jazz

8 Dinâmica - Variação de Amplitude

3.8.1 Definição Técnica

> **Feature:** Dynamic Range (variação de amplitude/volume)
> **Algoritmo:** Análise RMS (Root Mean Square) em escala dB
> **Implementação:**
> `rms = librosa.feature.rms(y=audio)`
> `db = librosa.amplitude_to_db(rms, ref=np.max)`
> `dynamic_range = max(db) - min(db)`
> **Medições:**
> • RMS: Amplitude média quadrática
> • dB: Escala logarítmica (20 × log10(RMS / ref))
> • Range: Diferença entre máximo e mínimo
> **Range típico:** 10-60 dB

3.8.2 O Que Significa (Linguagem Natural)

> **Analogia:** É quanto você varia o volume entre cantar FORTE e FRACO. Como ajustar o volume de um rádio constantemente vs. deixar fixo.
> **Por que importa:** A dinâmica controla a expressividade. Whitney Houston tem altíssimo range dinâmico (sussurra nos versos, grita nos refrões). Billie Eilish tem baixo range (volume quase constante, moderno).
> **Escala dB:**
> - 0 dB = Volume máximo (referência)
> - -10 dB = Forte
> - -30 dB = Médio
> - -50 dB = Fraco

3.8.3 Dados Extraídos

> **Output:**
> - `rms_media` : RMS médio
> - `rms_std` : Variabilidade de RMS
> - `db_media` : Volume médio (dB)

- `db_std` : Variabilidade de volume
- `db_min` : Volume mais baixo
- `db_max` : Volume mais alto
- `range_dinamico` : max - min (dB)

3.8.4 Interpretação

Range (dB)	Classificação	Característica	Exemplos
< 20	Muito Baixa	Volume quase constante, comprimido	Pop moderno produzido
20 - 35	Baixa	Pouca variação, controlado	Indie, Folk
35 - 45	Média	Variação natural	Pop, Rock
> 45	Alta	Contraste extremo (pp → ff)	Soul, Gospel, Clássico

9 Portamentos - Transições de Pitch

3.9.1 Definição Técnica

Feature: Portamento/Glissando (transições contínuas de pitch)
Algoritmo: Análise de gradiente em contorno de F0
Implementação:

```
diff = np.diff(f0_clean)
gradientes = np.gradient(diff)
# Portamento: 1 < |gradiente| < 10 por >5 frames
```

Critérios de detecção:

- Mudança gradual (não abrupta)
- 1 Hz/frame < mudança < 10 Hz/frame
- Duração mínima: 5 frames consecutivos

Output: Número de portamentos e duração média

3.9.2 O Que Significa (Linguagem Natural)

Analogia: É quando você "escorrega" de uma nota para outra ao invés de pular direto. Como subir uma escada (sem portamento) vs. um escorregador (com portamento).
Por que importa: Portamentos definem o estilo de articulação. R&B e Gospel usam MUITOS portamentos (melismático). Rock e Punk usam POUCOS (direto, staccato).
Som que faz:

- SEM: "Dó - Mi - Sol" (notas separadas)
- COM: "Dóóóó~Miiii~Sóóól" (desliza)

3.9.3 Dados Extraídos

Output:

- `portamentos_detectados` : Número total
- `duracao_media_segundos` : Duração média de cada

3.9.4 Interpretação

Quantidade	Classificação	Estilo de Articulação	Gêneros
< 5	Poucos/Ausentes	Direto, staccato, separado	Rock, Punk, Pop direto
5 - 20	Moderado	Algumas conexões legato	Pop, Country
> 20	Muitos	Melismático, escorregado	R&B, Gospel, Soul

10 HNR - Relação Harmônicos/Ruído

3.10.1 Definição Técnica

> **Feature:** Harmonics-to-Noise Ratio
> **Algoritmo:** Autocorrelação via Praat
> **Implementação:**
> ```
> snd = parselmouth.Sound(audio)
> harmonicity = snd.to_harmonicity_cc(0.01, 75, 0.1, 1.0)
> hnr = harmonicity.get_mean(0, 0)
> ```
> **Fórmula:** HNR = 10 × log10(E_harmônica / E_ruído)
> **Range:** 0 dB (muito rouquenho) a 30+ dB (muito limpo)
> **Interpretação clínica:** HNR < 10 dB pode indicar disfonia

3.10.2 O Que Significa (Linguagem Natural)

> **Analogia:** É quão "limpa" vs. "rouca" é sua voz. Como água mineral cristalina (alto HNR) vs. água com areia misturada (baixo HNR).
> **Por que importa:** HNR mede a "qualidade" da produção vocal. Alto HNR geralmente indica boa técnica e cordas vocais saudáveis. Baixo HNR pode ser estilo artístico (Louis Armstrong) ou problema de saúde vocal.
> **O que afeta o HNR:**
> - Hidratação das cordas vocais
> - Fechamento glótico (técnica)
> - Lesões nas cordas (nódulos, pólipos)
> - Estilo intencional (voz rouca como efeito)

3.10.3 Dados Extraídos

> **Output:**
> - `hnr` : Valor numérico em dB

3.10.4 Interpretação

HNR (dB)	Qualidade	Descrição	Observações
> 20	Excelente	Voz muito limpa, cristalina	Ótima técnica, cordas saudáveis
15 - 20	Muito Boa	Voz limpa, clara	Boa técnica
10 - 15	Boa	Voz natural, saudável	Normal
5 - 10	Rouca	Voz com ruído perceptível	Pode ser estilo ou problema inicial
< 5	Muito Rouca	Voz predominantemente ruidosa	Consultar otorrinolaringologista

> ⚠️ **NOTA CLÍNICA:** HNR persistentemente baixo (<10 dB) por semanas pode indicar lesão nas cordas vocais (nódulos, pólipos, edema). Este sistema não substitui avaliação médica. Consulte um otorrinolaringologista se HNR estiver consistentemente baixo.

11 Ressonância - Distribuição Espectral por Registro

3.11.1 Definição Técnica

Feature: Vocal Register (distribuição de energia por faixas de frequência)
Algoritmo: Análise espectral segmentada
Faixas analisadas:
• Peitoral: 80-350 Hz
• Mista: 350-600 Hz
• Cabeça: 600-1500 Hz
• Nasal: 1500-3000 Hz
• Falsete: 800-2000 Hz
Cálculo:

```
 for faixa in faixas:
  mask = (freqs >= f_min) & (freqs <= f_max)
  energia[faixa] = sum(magnitude[mask])
registro = max(energia, key=energia.get)
```

Threshold: Registro dominante tem >40% da energia total

3.11.2 O Que Significa (Linguagem Natural)

Analogia: É ONDE sua voz está "vibrando" - no peito, na cabeça, no nariz? Como diferentes caixas de ressonância de instrumentos musicais.
Por que importa: O registro define o caráter fundamental da voz. Voz de peito é "grossa", "potente". Voz de cabeça é "leve", "aguda". Voz mista mistura os dois.
Teste físico simples:

- Coloque a mão no peito e cante - vibra? = Voz de peito
- Toque o topo da cabeça e cante - vibra? = Voz de cabeça
- Tampe o nariz e cante - muda? = Ressonância nasal

3.11.3 Dados Extraídos

Output:

- `tipo` : Registro predominante (peitoral/mista/cabeça/nasal/falsete)
- `freq_dominante` : Frequência com maior energia (Hz)

- `marcacao` : Tag para uso no relatório ([PEITO], [CABEÇA], etc.)

3.11.4 Interpretação por Registro

Registro	Faixa (Hz)	Sensação Física	Característica Sonora
Peitoral	80-350	Vibração no peito	Grave, potente, "grosso"
Mista	350-600	Peito + cabeça simultâneos	Transição, equilibrado
Cabeça	600-1500	Vibração no crânio	Agudo, leve, "fino"
Nasal	1500-3000	Vibração no nariz/face	Anasalado (country, alguns estilos)
Falsete	800-2000	Leve, sem vibração peitoral	Muito agudo, aéreo, leve

3.11.5 Ranges Típicos por Gênero

Gênero	Voz de Peito	Voz Mista	Voz de Cabeça/Falsete
Masculino	E2 - E4 (80-330 Hz)	E4 - A4 (330-440 Hz)	A4+ (440+ Hz)
Feminino	A2 - E4 (110-330 Hz)	E4 - C5 (330-520 Hz)	C5+ (520+ Hz)

4. Implementação e Uso

4.1 Instalação no Google Colab

Passo 1: Acesso

1. Acesse: `colab.research.google.com`
2. Faça login com conta Google
3. Crie novo notebook ou faça upload do arquivo `.ipynb`

Passo 2: Montar Google Drive

```
from google.colab import drive
drive.mount('/content/drive')
```

Passo 3: Configurar Caminhos

```
TRACK = "Nome_Artista_Musica"
AUDIO_PATH = '/content/drive/MyDrive/pasta/arquivo_vocals.wav'
MODEL_SIZE = 'medium'  # tiny, base, small, medium, large
SEGMENTACAO_TEMPO = 3.0  # segundos por segmento
```

4.2 Parâmetros Configuráveis

4.2.1 Modelo Whisper

Modelo	Velocidade	Precisão	Uso RAM	Recomendação
tiny	Muito rápido (1-2 min)	~70%	1 GB	Teste rápido apenas
base	Rápido (2-3 min)	~80%	1.5 GB	Preview inicial
small	Médio (3-5 min)	~85%	2 GB	Bom custo-benefício
medium	Lento (5-10 min)	**~95%**	5 GB	**RECOMENDADO**

Modelo	Velocidade	Precisão	Uso RAM	Recomendação
large	Muito lento (15-25 min)	~98%	10 GB	Máxima precisão

4.2.2 Segmentação Temporal

Tempo (s)	Granularidade	Segmentos Típicos (3 min)	Uso Ideal
1.0	Palavra por palavra	~180	Análise extremamente detalhada
3.0	Frase curta	**~60**	**RECOMENDADO - melhor balanço**
5.0	Frase longa	~36	Visão geral de frases completas
10.0	Seção musical	~18	Análise de verso/refrão/ponte

4.3 Execução

> **Método 1: Célula por Célula**
>
> 1. Clique na primeira célula
> 2. Pressione **Shift + Enter**
> 3. Aguarde execução
> 4. Repita para cada célula
>
> **Método 2: Executar Tudo (RECOMENDADO)**
>
> 1. Menu: **Runtime → Run all**
> 2. Ou pressione **Ctrl + F9**
> 3. Aguarde 10-20 minutos

4.4 Progresso de Execução

> `Mensagens esperadas durante execução:`

```
[1/5] Carregando modelo Whisper (medium)...
[2/5] Detectando idioma... ✓ Idioma: PT/EN/ES/etc
[3/5] Transcrevendo áudio... ✓ X segmentos detectados
[4/5] Aplicando segmentação de 3.0s... ✓ Y segmentos após refinamento
[5/5] Analisando características vocais...
  Analisando segmento 5/Y...
  Analisando segmento 10/Y...
  ...
✓ Análise completa!
✓ Arquivo salvo: /caminho/arquivo_analise_pratica.txt
```

5. Interpretação de Resultados

5.1 Estrutura do Relatório

> O relatório gerado tem duas partes principais:
>
> **PARTE 1: Guia Prático Global**
>
> Análise das características médias de toda a música, com instruções práticas em linguagem natural para cada feature.
>
> **PARTE 2: Análise Técnica por Segmento**
>
> Dados detalhados para cada frase/segmento da música, permitindo análise verso-a-verso.

5.2 Exemplo de Output - Parte 1 (Guia Prático)

```
================================================================================
GUIA PRÁTICO DE REPLICAÇÃO VOCAL - O QUE FAZER PARA IMITAR ESTA VOZ
================================================================================

1. ALTURA DA VOZ (PITCH/TOM) - Qual nota cantar
--------------------------------------------------------------------------------

📊 DADOS TÉCNICOS:
   • F0 Médio: 464.3 Hz
   • Range: 350.0 Hz a 580.0 Hz

🎯 O QUE FAZER:
   ✓ Esta é uma VOZ MUITO AGUDA
   ✓ Cante acima de F4 (FÃ¡4)
   ✓ Use FALSETE ou voz de cabeça pura
   ✓ Som leve e aéreo

💡 COMO PRATICAR:
   1. Use um afinador de celular (apps: Tuner, Pano Tuner)
   2. Cante uma nota sustentada tentando manter 464 Hz
   3. Ajuste até o afinador mostrar a frequência correta
```

```
2. VIBRATO - A 'tremidinha' na voz
──────────────────────────────────

📊 DADOS TÉCNICOS:
   • Presente: SIM
   • Taxa: 3.7 Hz (oscilações por segundo)
   • Extensão: 817.8 cents

🎯 O QUE FAZER:
   ✓ Vibrato LENTO: ondulações lentas e amplas
   ✓ Faça cerca de 4 oscilações por segundo
   ✓ Vibrato INTENSO: variação grande e expressiva

[... continua com todos os 11 features ...]
```

5.3 Exemplo de Output - Parte 2 (Análise por Segmento)

```
================================================================================
ANÁLISE TÉCNICA DETALHADA POR SEGMENTO
================================================================================
Segmentação: 3.0s por segmento
Total de segmentos: 71
================================================================================

[0.50s - 3.50s] "I'm gonna swing from the chandelier"
──────────────────────────────────────────────────────────────────────────────
1️⃣ PITCH: 464.3 Hz (MUITO AGUDO) | Range: 350-580 Hz
2️⃣ VIBRATO: 3.7 Hz (LENTO), 817.8 cents (intenso)
3️⃣ FORMANTES: F1=850Hz (boca ABERTA, mandíbula baixa) | F2=1800Hz (língua CENTRO,
4️⃣ BREATHINESS: ALTO (ratio: 0.452) → deixe MUITO ar passar, cante sussurrado
5️⃣ BRILHO: BRILHANTE (2800 Hz) → som para FRENTE, máscara facial
6️⃣ TENSÃO: MÉDIA (ratio: 0.185) → tensão equilibrada
7️⃣ ATAQUE: 0.089s (MODERADO) → comece firme mas não abrupto
8️⃣ DINÂMICA: -15.2 dB (FORTE) | Range: 35.8 dB (variação moderada)
9️⃣ PORTAMENTOS: 12 detectados (alguns deslizes) → deslize ocasionalmente
🔟 QUALIDADE: HNR 16.3 dB (voz MUITO LIMPA) → som cristalino, pouco ruído
➕ RESSONÂNCIA: MISTA (580 Hz) - transição entre registros

[3.50s - 6.50s] "From the chandelier"
──────────────────────────────────────────────────────────────────────────────
```

```
[... continua para todos os segmentos ...]
```

5.4 Como Usar os Dados

Estratégia de Aprendizado Progressivo:

Semana 1: Pitch (Altura)

- Foque APENAS em acertar a altura média
- Use afinador constantemente
- 15 minutos/dia de prática

Semana 2: Pitch + Vibrato

- Mantenha o pitch correto
- Adicione as oscilações do vibrato
- Use metrônomo para controlar taxa

Semana 3-4: Formantes

- Ajuste abertura da boca (F1)
- Mova posição da língua (F2)
- Grave e compare com original

Semana 5-6: Características Secundárias

- Breathiness, brilho, tensão
- Um elemento novo a cada 3 dias

Semana 7-8: Integração

- Todos os elementos simultâneos
- Grave sua versão final
- Analise SEU áudio com o sistema
- Compare números com o original

6. Limitações e Considerações

6.1 Limitações Técnicas

1. Separação de Fonte (Demucs)

Problema: Pode deixar "vazamento" de instrumentos nos vocais isolados
Impacto: Afeta principalmente formantes, HNR e breathiness
Solução: Use vocais já isolados profissionalmente quando possível
Mitigação: Resultados ainda são 80-90% precisos na maioria dos casos

2. Técnicas Vocais Extremas

Problema: Growl, scream, fry, death metal vocals não são bem capturados
Motivo: Algoritmos otimizados para voz cantada "normal"
Limitação: F0 pode não ser detectado; formantes imprecisos
Funciona parcialmente: Dinâmica, ataque e ressonância ainda são úteis

3. Harmonias Vocais Simultâneas

Problema: Múltiplas vozes ao mesmo tempo confundem os algoritmos
Impacto: F0, formantes e vibrato ficam imprecisos
Solução: Analise apenas lead vocals isolados

4. Qualidade de Gravação

Formato aceito: Apenas arquivos WAV
Qualidade recomendada: 44.1 kHz, 16-bit ou superior
Problema com qualidade inferior: Áudios de baixa resolução podem afetar precisão da análise espectral, especialmente formantes e HNR
Conversão de outros formatos: Se você tem MP3, FLAC ou outros formatos, converta para WAV antes de processar (use Audacity, FFmpeg ou conversores online)

6.2 Limitações de Interpretação

O Sistema NÃO Mede:

- Postura corporal (que afeta produção vocal)
- Tensão muscular extralaríngea
- Expressão facial
- Aspectos subjetivos ("emoção", "feeling")
- Contexto cultural/estilístico

6.3 Considerações de Saúde Vocal

⚠️ **IMPORTANTE:** Este sistema é uma ferramenta de análise, NÃO um substituto para:

- Aulas de canto com professor qualificado
- Avaliação médica (otorrinolaringologista)
- Tratamento fonoaudiológico

Sinais de alerta (procure médico):

- HNR consistentemente <8 dB por mais de 2 semanas
- Dor ao cantar
- Rouquidão persistente
- Perda de range vocal
- Fadiga vocal excessiva

6.4 Uso Responsável

Recomendações:

- Use os dados como GUIA, não como regra absoluta
- Respeite os limites naturais da sua voz
- Não force técnicas que causem desconforto
- Progressão gradual: 15-30 min/dia é suficiente

- Hidratação adequada (2-3 litros água/dia)
- Descanso vocal adequado

7. Referências Técnicas

7.1 Algoritmos e Métodos

Pitch (F0) Detection:

• De Cheveigné, A., & Kawahara, H. (2002). "YIN, a fundamental frequency estimator for speech and music". Journal of the Acoustical Society of America, 111(4), 1917-1930.

Formant Analysis:

• Burg, J. P. (1975). "Maximum entropy spectral analysis". Ph.D. dissertation, Stanford University.
• Boersma, P., & Weenink, D. (2021). "Praat: doing phonetics by computer" [Software]. Version 6.1.38.

Voice Quality (HNR, Jitter, Shimmer):

• Boersma, P. (1993). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound". IFA Proceedings, 17, 97-110.

Spectral Features:

• McFee, B., et al. (2015). "librosa: Audio and music signal analysis in python". Proceedings of the 14th Python in Science Conference.

Source Separation:

• Défossez, A., et al. (2019). "Demucs: Deep Extractor for Music Sources". arXiv:1909.01174.

Speech Recognition:

• Radford, A., et al. (2022). "Robust Speech Recognition via Large-Scale Weak Supervision". arXiv:2212.04356.

7.2 Código-Fonte e Documentação

Repositório GitHub: [URL do repositório]
Licença: MIT License (código aberto)
Documentação completa: README.md no repositório

Issues/Bugs: GitHub Issues

Contribuições: Pull requests são bem-vindos

7.3 Leitura Adicional Recomendada

Fundamentos de Acústica Vocal:

• Sundberg, J. (1987). "The Science of the Singing Voice". Northern Illinois University Press.

Processamento de Sinais de Áudio:

• Smith, J. O. (2011). "Spectral Audio Signal Processing". W3K Publishing.

Análise Vocal Computacional:

• Kreiman, J., & Sidtis, D. (2011). "Foundations of Voice Studies". Wiley-Blackwell.

Informações de Contato

Suporte Técnico: Issues no GitHub
Documentação: README.md no repositório
Comunidade: Discussions no GitHub

Desenvolvido para pesquisa e educação em técnica vocal
Este sistema não substitui orientação profissional médica ou pedagógica

Versão da Documentação: 5.0
Última Atualização: Janeiro 2025
Compatibilidade: Google Colab, Python 3.8+