VÍDEO DE LANÇAMENTO · VOZ, TRILHA E SOM

Voz, trilha e som de vídeo imobiliário: o checklist com os números que usamos

Por Ben-Hur Real Figueiró, diretor de fotografia e AI engineer, fundador da 10Dobro Prod.

O que mudou do off reprovado para o aprovado num filme de lançamento em São Luís: direção de voz, tags de emoção, a régua de medição, a trilha escolhida pela emoção e o mix.



RESUMO

- Voz de vídeo imobiliário funciona quando conta em vez de ler: direção em uma linha, uma tag de emoção por frase e três takes do off inteiro.
- A régua que usamos: 3 a 6 semitons de variação de altura e 3 a 4,5 sílabas por segundo. Abaixo de 2,5 semitons é leitura plana; se não cabe no tempo, corta texto.

> - A trilha sai da emoção do arco, não da vontade de variar: piano e cordas a 82 bpm no nosso filme, retirada no vale e entrada cheia na virada.
> - No mix, a trilha fica cerca de 20 dB abaixo da voz enquanto ela fala, -14 LUFS para redes, -16 para site e pico abaixo de -1 dBTP.

A direção que mudou a voz do nosso filme

No filme de 36 segundos de um lançamento em São Luís, a primeira versão do off foi reprovada: "sem emoção". Medimos e o motivo apareceu nos números. A voz variava só 1,2 a 2,9 semitons e corria a 5,2 a 6,4 sílabas por segundo, porque tinha texto demais para o tempo.

A segunda versão começou por uma linha de direção, escrita antes de gerar qualquer áudio:

```
Uma amiga contando pra outra, com sorriso na voz, íntima, sem vender.
```

As tags que funcionaram e as que achataram

Na ElevenLabs v4, a tag entre colchetes no começo da frase espalha o tom pela frase inteira. Fizemos uma bateria de 12 variações com a mesma voz antes de escolher.

FUNCIONARAM	ACHATARAM
[warmly], **[smiling]**, **[curious, warm]**, **[happy, warm]**	[softly], [whispering], [dreamy], [in awe]
entraram na régua de 3 a 6 semitons	1,9 a 2,4 semitons de variação, e a fala acelerou

> **Por que as suaves falham.** Sussurro e encanto parecem combinar com casa e família, mas tiram a variação de altura da voz. O resultado soa plano, não íntimo.

Abaixo, liberado com o cadastro: a régua de medição, as 7 falas do filme com tag e números, a escolha da trilha (e o erro da primeira), o desenho de som, os números do mix e

um checklist para imprimir.

1. A régua de medição da voz

Quem monta o vídeo nem sempre consegue ouvir com calma cada take. Por isso a voz passa por número antes de passar pelo ouvido de quem aprova.

MEDIDA	FAIXA	O QUE SIGNIFICA
Variação de altura	3 a 6 semitons	narração com emoção. De 2,5 a 3 é limítrofe; abaixo de 2,5, leitura plana: descartar
Velocidade	3 a 4,5 sílabas/s	tom íntimo. Até 5 para tom dinâmico; acima de 5 soa corrido
Folga	0,3 s	a frase cabe na janela do plano com folga

- **Três takes do off inteiro, numa chamada só.** Gerar frase por frase quebra a continuidade de timbre e energia entre os trechos. Depois corte as frases pelos tempos de cada palavra (endpoint com timestamps).
- **Refazer uma frase só:** mande junto o texto vizinho (antes e depois) para a entonação continuar a mesma.
- **Não cabe? Corte texto.** No modelo v4 o parâmetro de velocidade não muda a fala. E voz acelerada é justamente o que reprova.
- **Os takes escolhidos do nosso filme** ficaram entre 3,73 e 4,37 semitons e entre 3,8 e 4,6 sílabas por segundo.

2. Roteiro para o ouvido

- Frases curtas, de até 12 palavras. Uma ideia por frase.
- Reticências no texto da voz marcam pausa. A legenda leva a pontuação normal.
- Número por extenso no texto da voz ("trezentos metros") e em algarismo na tela ("300 metros"). Metragem e planta ficam só na tela.
- Respiro entre as frases: no nosso filme, as 7 falas somam 26,6 s de voz num filme de 35,8 s.

As 7 falas do filme, com o nome do empreendimento trocado por marcador:

FALA	LEGENDA	TEXTO PARA A VOZ	TAKE ESCOLHIDO
F1	Daqui de cima, dá pra ver tudo o que fica perto.	[warmly] Daqui de cima... dá pra ver tudo o que fica perto.	3,77 st · 4,21 síl/s
F2	No [bairro], em São Luís, nasce o [nome do empreendimento].	[warmly, proud] No [bairro], em São Luís... nasce o [nome do empreendimento].	3,73 st · 4,28 síl/s
F3	O mercado fica a 300 metros. E a escola, a 1 minuto.	[warmly] O mercado fica a trezentos metros. E a escola... a um minuto.	3,76 st · 3,80 síl/s
F4	O fim de semana começa sem sair de casa.	[happy, playful] O fim de semana... começa sem sair de casa.	4,22 st · 4,56 síl/s
F5	E lá dentro, o dia inteiro cabe.	[warmly, amazed] E lá dentro... o dia inteiro cabe.	4,37 st · 4,28 síl/s
F6	Da luz da manhã ao fim de tarde na varanda.	[smiling] Da luz da manhã... ao fim de tarde na varanda.	3,77 st · 4,31 síl/s
F7	Aqui, o seu sonho tem endereço.	[happy, warm] Aqui, o seu sonho... tem endereço.	4,35 st · 4,12 síl/s

As falas 3 e 7 precisaram de segunda e terceira rodadas: a 3 saía corrida e a 7 saía plana. O off inteiro, pronto para colar:

```
[warmly] Daqui de cima... dá pra ver tudo o que fica perto.
[warmly, proud] No [bairro], em São Luís... nasce o [nome do empreendimento].
[warmly] O mercado fica a trezentos metros. E a escola... a um minuto.
[happy, playful] O fim de semana... começa sem sair de casa.
[warmly, amazed] E lá dentro... o dia inteiro cabe.
[smiling] Da luz da manhã... ao fim de tarde na varanda.
[happy, warm] Aqui, o seu sonho... tem endereço.
```

3. Trilha escolhida pela emoção, e o erro da primeira

Na primeira versão a trilha foi Electro House a 128 bpm. O motivo da escolha: era um estilo que a gente nunca tinha usado. Variar não é critério. A pergunta certa é qual emoção o arco pede em cada trecho.

TOM DA PEÇA	INSTRUMENTAÇÃO	ANDAMENTO
família, casa, conquista	piano, cordas, violão, pad suave, percussão leve só no clímax	70 a 100 bpm
institucional cinematográfico	cordas, pad, piano grave, percussão orquestral	80 a 100 bpm
tecnologia e dados	pulso eletrônico limpo, arpejo, baixo sintetizado	100 a 124 bpm
bastidor leve, Reels de dica	groove acústico ou eletrônico leve	90 a 115 bpm
tensão, erro, mistério	drones, pulsos graves, silêncio	livre

A trilha da segunda versão: original, montada com Apple Loops, em Dó maior a 82 bpm, com piano e cordas (família e conquista). A 82 bpm cada batida dura 0,732 s, e os cortes principais caem em tempo forte.

TRECHO DO FILME	O QUE A TRILHA FAZ
abertura	piano sozinho
maquete subindo	entram os plucks
localização	entra o baixo
lazer	cordas e metais
vale (a planta se desenhando)	só o piano, a -17 dB, por cerca de 2 s sem voz
virada (a planta vira a sala)	volta cheia no tempo forte
pico (varanda no fim de tarde)	pandeiro só aqui, trilha +4,5 dB, 1,5 s sem voz
fim	acorde aberto e cauda natural, sem corte seco

Levamos uma **trilha B** para escuta: Lá menor a 70 bpm, com violão, piano elétrico e pad. Mais íntima e nostálgica, mas fora da grade de cortes. Quem aprova escolhe de ouvido entre duas ou três candidatas boas, cada uma com uma linha dizendo por que serve ao arco.

4. Licença da trilha

- **Apple Loops:** licença Apple, sem crédito obrigatório. Vêm com o GarageBand e o Logic.
- **CC0:** livre, sem crédito obrigatório.
- **CC BY:** com crédito na legenda e no texto do anúncio: "Música: título, autor (CC BY versão)".
- **Nunca:** BY-SA, NC ou ND. E nada de música gerada por IA.

5. Som em camadas

De baixo para cima: ambiência, ruído da ação (passos, porta, louça, tecido, água), transições, impactos discretos, trilha e voz.

- **Em todo plano, uma ambiência e pelo menos um evento sincronizado com a ação.** No nosso filme: vento e cidade no alto, piscina e crianças longe, chute na bola, latido, papel e lápis no vale, pássaros da manhã, tecido na suíte.
- **J-cut:** o som do próximo plano entra antes da imagem. A água da piscina chega antes do corte; a risada de criança vem de outro cômodo.
- **Som sem ação na imagem sai.** Depois da crítica tiramos a xícara, as taças e a persiana, que tocavam sem nada acontecendo na tela.
- **Silêncio é ferramenta:** o vale antes da virada pede tirar camadas, não acrescentar.
- **Fontes:** Freesound CC0, escolhido cena a cena (nunca sorteio), e a biblioteca de efeitos do Final Cut Pro.

6. Mix: os números do filme

- A trilha abaixa sob a voz pelo envelope da própria voz: -22 dB enquanto ela fala (-12 dB no pico).
- Rampas de descida de cerca de 120 ms e de volta de 350 ms; pausa curta (menos de 0,9 s) mantém a trilha abaixada, para não bombear.
- Medido em LUFS só nos trechos de fala: trilha 19,9 dB e ambiência 22,8 dB abaixo da voz.
- Voz com passa-alta, compressor suave e presença; a trilha ganhou um recorte fixo em 1,6 e 2,5 kHz para não brigar com a fala.

ARQUIVO	INTEGRADO	LRA	PICO REAL
master (site)	-16,0 LUFS	4,5 LU	-1,7 dBTP
redes (Reels)	-14,1 LUFS	4,1 LU	-1,7 dBTP
versão com a trilha B	-16,0 LUFS	2,8 LU	-1,9 dBTP

> **LRA de 2,5 para 4,5 LU.** Na primeira passada da segunda versão o filme tinha faixa dinâmica de 2,5 LU e soava achatado. Depois da crítica, a trilha subiu no pico e o LRA foi a 4,5 LU. Medida momentânea: lazer -16,6, vale -29,2, pico sem voz -13,0 LUFS.

Alvos: -14 LUFS para redes e -16 para site, pico real abaixo de -1 dBTP. Meça com `ffmpeg -af ebur128`.

7. Custo

- **Voz:** ElevenLabs em caracteres da assinatura, não por vídeo. Na segunda versão foram cerca de 45 chamadas entre takes e baterias de teste. Confira no seu plano o uso comercial de vozes da biblioteca.
- **Trilha:** Apple Loops vêm com o GarageBand e o Logic.
- **Efeitos:** CC0 do Freesound e biblioteca do Final Cut Pro.

8. Checklist para imprimir

- [ ] Direção da voz escrita em uma linha: quem fala, para quem, em que situação.
- [ ] Frases com até 12 palavras, reticências onde quer pausa, números por extenso no texto da voz.
- [ ] Uma tag de emoção ativa por frase; nada de [softly], [whispering], [dreamy] ou [in awe] para narração calorosa.
- [ ] Off inteiro gerado numa chamada só, três takes.
- [ ] Cada take medido: 3 a 6 semitons, 3 a 4,5 sílabas por segundo, pelo menos 0,3 s de folga na janela do plano.
- [ ] Os dois melhores takes de cada frase ouvidos por quem aprova.
- [ ] Trilha escolhida pela emoção do arco, com licença Apple Loops, CC0 ou CC BY (crédito na legenda).
- [ ] Trilha retirada no vale, entrada cheia na virada, cauda no fim, cortes no compasso.
- [ ] Ambiência em todo plano e um evento sincronizado com a ação; nenhum som de ação que não está na imagem.

- [ ] Trilha 15 a 20 dB abaixo da voz enquanto ela fala; ambiência 20 a 28 dB abaixo.
- [ ] Medido com ebur128: -14 LUFS redes, -16 LUFS site, pico abaixo de -1 dBTP, LRA acima de 4 LU.

9. Passo a passo resumido

1. **Escreva a direção da voz em uma linha.** Quem fala, para quem e em que situação. No nosso filme: uma amiga contando pra outra, com sorriso na voz, sem vender.
2. **Escreva o roteiro para o ouvido.** Frases curtas, reticências onde quer pausa, números por extenso no texto da voz e na grafia normal na tela.
3. **Gere três takes do off inteiro.** Uma chamada com o texto todo, três vezes. Depois corte as frases pelos tempos de cada palavra.
4. **Meça cada take.** Variação de altura entre 3 e 6 semitons e 3 a 4,5 sílabas por segundo para tom íntimo. Abaixo de 2,5 semitons é leitura plana.
5. **Escolha a trilha pela emoção do arco.** Família e casa pedem piano, cordas ou violão a 70 a 100 bpm. Corte no compasso e deixe cauda no fim.
6. **Monte o som em camadas.** Ambiência em todo plano, um evento sincronizado com a ação e silêncio no vale antes da virada.
7. **Mixe e meça.** Trilha 15 a 20 dB abaixo da voz enquanto ela fala, -14 LUFS para redes, -16 para site, pico abaixo de -1 dBTP.

Perguntas frequentes

Que voz usar num vídeo de lançamento imobiliário?

Uma voz que conte em vez de ler. No nosso filme usamos uma voz feminina da biblioteca da ElevenLabs no modelo v4, dirigida como uma amiga contando pra outra, com sorriso na voz e sem vender. Antes de usar voz da biblioteca em anúncio, confira no seu plano da ElevenLabs as regras de uso comercial.

Quais tags de emoção funcionaram na ElevenLabs v4?

[warmly], [smiling], [curious, warm] e [happy, warm]. As tags suaves [softly], [whispering], [dreamy] e [in awe] achataram a voz para 1,9 a 2,4 semitons de variação e aceleraram a fala, numa bateria de 12 variações.

Quantos takes gravar?

Três takes do off inteiro, numa chamada só, e não frase por frase. Gerar frase por frase quebra a continuidade de timbre e entonação. Depois meça cada take e leve os dois melhores de cada frase para ouvir.

A voz ficou corrida. Aumento ou diminuo a velocidade?

Corte texto. No modelo v4 o parâmetro de velocidade não muda a fala. A primeira versão do nosso off, reprovada, media 5,2 a 6,4 sílabas por segundo porque tinha texto demais para o tempo. Os takes aprovados ficaram entre 3,8 e 4,6.

Posso usar música do Instagram no anúncio?

Para anúncio e para o site, use trilha com licença clara. Nós usamos trilha original montada com Apple Loops, que não pede crédito, ou faixa CC0 ou CC BY com crédito na legenda. Licenças BY-SA, NC e ND ficam de fora, e não usamos música gerada por IA.

Preciso dar crédito à trilha?

Depende da licença. Apple Loops não pedem crédito. CC BY pede crédito com título, autor e licença. CC0 não exige, mas dar crédito não custa nada.

Qual volume usar para Reels?

-14 LUFS integrado com pico real abaixo de -1 dBTP. Para o site, -16 LUFS. O nosso filme saiu com -14,1 LUFS na versão de redes e -16,0 no master, os dois com pico de -1,7 dBTP.

Quanto custa a voz e a trilha?

A voz da ElevenLabs sai em caracteres da assinatura, não em cobrança por vídeo. Os Apple Loops vêm com o GarageBand e o Logic. Os efeitos que usamos são CC0 do Freesound e da biblioteca do Final Cut Pro.

Os bastidores dos nossos filmes continuam no Instagram: @10dobroprod. Outros materiais grátis em Método 10X.

Por Ben-Hur Real Figueiró, diretor de fotografia e AI engineer, fundador da 10Dobro Prod. Melhor Direção de Fotografia no Festival Guarnicê 2024.

10DOBROPROD.COM.BR · SÃO LUÍS, MA