SoftGestos: comunicar com pessoas surdas sem sair do telemóvel

Publicado em 2026-11-27 13:55:04



Há conversas que começam com um gesto e acabam num encolher de ombros, porque um dos lados não sabe a língua do outro. O SoftGestos nasceu para encurtar essa distância, sem pedir nada a ninguém: nem conta, nem internet, nem um servidor algures a ouvir.

Este é o relato técnico de como a app foi pensada, do que faz bem, do que ainda não faz, e de porque é que tudo acontece dentro do telemóvel.

[LINK DE DOWNLOAD]

O problema

Quem ouve e quem não ouve partilham o mesmo café, a mesma fila do balcão, o mesmo consultório. Falta-lhes muitas vezes uma ponte rápida para o essencial: um nome, um pedido, um «obrigado». As soluções que existem tendem a depender da nuvem, o que significa enviar imagens da cara e das mãos de alguém para servidores de terceiros. Para uma ferramenta que filma pessoas, isso pareceu-me inaceitável.

A regra de partida foi simples: **tudo no aparelho**. A câmara, o reconhecimento das mãos, a voz, os dados. A app nem sequer pede permissão de Internet.

O que a app faz

- **Gestos → texto e voz.** A câmara segue as mãos e reconhece a datilologia da língua gestual americana (ASL), letra a letra, e gestos isolados de palavras. O texto aparece no ecrã e pode ser dito em voz alta.

ou não pode falar escreve, e o telemóvel diz.

- **Gestos pessoais.** Qualquer gesto pode ser gravado com 3 a 5 amostras e passa a ser reconhecido. Isto vale para qualquer língua gestual, incluindo a **LGP**.
- **Conversas guardadas cifradas**, só quando o utilizador carrega em Guardar.

Porquê ASL e não LGP? Porque praticamente não há dados abertos de língua gestual portuguesa para treinar seja o que for. A ASL tem conjuntos públicos razoáveis. Os gestos pessoais são a forma honesta de abrir a porta à LGP: cada pessoa ensina à app os gestos que usa.

Por dentro: da câmara à letra

A cadeia é curta e corre toda em tempo real:

1. **CameraX** entrega imagens de cerca de 640×480, guardando só a mais recente (se o processamento atrasar, as antigas são descartadas).
2. **MediaPipe Hand Landmarker** (modelo `hand_landmarker.task`, Apache 2.0) encontra até duas mãos e devolve 21 pontos 3D por mão. A imagem é rodada e espelhada para parecer sempre uma selfie, seja qual for a câmara.
3. O **motor**, escrito em Kotlin puro e testado numa JVM, normaliza os pontos, decide a letra ou o gesto e constrói o texto.

unidades iguais nos dois eixos. Depois a mão é centrada no pulso e dividida pelo seu próprio tamanho (distância do pulso à base do dedo médio), para que uma mão perto ou longe da câmara dê a mesma forma. Para as letras há ainda uma versão canónica: espelhada para parecer sempre mão direita e rodada para o dedo médio apontar para cima.

Letras por regras, sem treino

As letras estáticas do alfabeto ASL (A a Y, sem J e Z, que exigem movimento) são reconhecidas por regras, não por uma rede neuronal. Para cada dedo mede-se a retidão (distância da base à ponta a dividir pela soma dos três ossos), a posição do polegar face às articulações, o afastamento entre dedos, se o indicador cruza o médio, e o ângulo da palma na imagem, que é o que separa pares como U/H, D/G ou K/P.

Cada letra é uma conjunção difusa de condições com rampas lineares. Ganha a melhor, se passar a confiança mínima e tiver uma margem de pelo menos 0,1 sobre a segunda. Um estabilizador evita letras repetidas: a letra só se escreve depois de alguns fotogramas iguais, e as letras duplas (o «LL» de HELLO) pedem um pequeno deslize lateral. Baixar a mão durante 1,2 s dá um espaço.

prioridade.

Palavras: DTW e um modelo opcional

Uma palavra gestual é movimento, não forma. Um segmentador decide quando o gesto começa (mãos à vista) e quando acaba (mãos fora da imagem, 600 ms paradas, ou 4 s no máximo). A sequência é reamostrada para 32 fotogramas, cada um com 132 números (duas mãos × presença, posição do pulso e 63 coordenadas de forma).

Os gestos pessoais comparam-se com **DTW** (dynamic time warping) com banda de Sakoe-Chiba, que tolera que o mesmo gesto seja feito mais depressa ou mais devagar. É barato, não precisa de treino e chega para algumas dezenas de gestos.

Para centenas de gestos há um caminho opcional: um modelo **TensorFlow Lite** com entrada `[1, 32, 132]`. A arquitetura é pequena (uma camada densa, três blocos Conv1D residuais com LayerNorm, média temporal, softmax) e o ficheiro final ronda os 0,6 MB. Os gestos pessoais com confiança suficiente ganham sempre ao modelo.

cuidado, e não quis redistribuir nada cujo uso não estivesse claro. O projeto inclui os scripts para cada um treinar o seu, por exemplo com os dados públicos «ASL Signs» do Kaggle (250 gestos):

```
bash training/build_model.sh
```

O script precisa de uma conta Kaggle gratuita e do token da API. Descarrega os dados, prepara-os, treina e copia `signs.tflite` e `labels.txt` para a app. Também é possível importar os dois ficheiros diretamente no telemóvel, em «Os meus gestos → Importar modelo», ou juntar vídeos próprios ao treino.

Um detalhe de que me orgulho: a matemática das características existe duas vezes, em Kotlin (na app) e em Python (no treino). Um teste escreve um ficheiro de referência do lado Kotlin e outro confirma do lado Python que a diferença é zero. Se o modelo for treinado com números ligeiramente diferentes dos que a app calcula, falha em silêncio; este teste existe para isso nunca acontecer.

Os dados ficam num ficheiro JSON cifrado em vez de numa base de dados SQLite/Room. É um desvio consciente: para o volume de dados desta app, um ficheiro cifrado por inteiro é mais simples e não deixa nada em claro.

Uma nota sobre a voz: por omissão só se usa o reconhecimento no aparelho (Android 12 ou superior), e o áudio não sai do telemóvel. Se o aparelho não o tiver, o utilizador pode permitir o reconhecedor do sistema, que

- Kotlin 2.0.21, Jetpack Compose com Material 3 (BOM 2025.04.01)
- Android 10 (API 29) ou superior, compilado para a API 36; telemóvel e tablet na mesma app
- CameraX, MediaPipe Tasks Vision 0.10.20, TensorFlow Lite 2.17 (LiteRT)
- Interface bilingue, português e inglês, segundo o idioma do sistema
- Documentação de utilizador e técnica dentro da app, nas duas línguas

Limites, ditos às claras

> O SoftGestos não é um intérprete. É uma ajuda à comunicação.

- Não traduz frases contínuas; reconhece letras e gestos isolados.
- Expressões faciais e gramática da ASL não são interpretadas.
- A precisão das letras foi testada com mãos sintéticas (rotação, escala, ruído, mão esquerda), não medida com pessoas reais.
- Sem um modelo treinado, as palavras dependem dos gestos pessoais gravados.

Há algo de teimoso em fazer tudo localmente quando a moda é mandar tudo para a nuvem. Mas uma app que aponta uma câmara a uma pessoa tem de merecer a confiança dessa pessoa, e a forma mais simples de a merecer é não ter para onde enviar nada. O resto (mais gestos, melhor precisão, talvez um dia dados abertos de LGP) constrói-se em cima dessa base.

Se experimentar, diga-me o que falhou. É assim que as regras melhoram.

✦ AUTORIA E INVESTIGAÇÃO

*Artigo da autoria de **Francisco Gonçalves** e co-autoria de **Augustus Veritas**, na investigação de fontes.*

Fragmentos do Caos · Redação Editorial

⬇ Descarregar o SoftGestos

[LINK DE DOWNLOAD]

Blogue Fragmentos do Caos

A verdade nasce onde o pensamento é livre.

☁️ GitHub Pages

🛰️ CodeBerg Pages

🌌 **Fragmentos do Caos:** Blogue • Ebooks • Carrossel

👁️ Esta página foi visitada ... vezes.

Contactos