

IMPLEMENTAÇÃO DE UM SISTEMA DE CAPTURA DE MOVIMENTOS DE SINAIS DE LIBRAS PARA ANIMAÇÃO DE AVATARES 3D.

Aluno: Gabriel Junqueira do Val - RA: 211023558

Orientador: Prof. Assoc. Antonio Carlos Sementille.

1. Introdução
2. Fundamentação Teórica
3. Metodologia
4. Experimentos
5. Conclusões Finais

1. Introdução

- **A comunicação é um processo essencial da vida humana, de acordo com Cristiano (2017), ela pode se dar além da forma oral;**
 - Gestos, expressões faciais e corporais, articulações das mãos e braços fazem parte desse processo.
- **Pessoas que sofrem com algum tipo de deficiência auditiva ou oral, têm o seu primeiro contato com a comunicação a partir da língua de sinais do local onde vivem.**
 - LIBRAS - Língua Brasileira de Sinais;
 - ASL - *American Sign Language*;
 - LGP - Língua Gestual Portuguesa.

- **A LIBRAS é assegurada como forma válida de comunicação pela lei Nº10.436 de 24 de abril de 2002. (BRASIL, 2002). Sendo então, não uma tradução do português brasileiro, e sim uma língua própria;**
- **De acordo com Capovilla (2017), existem mais de 14 mil sinais de LIBRAS;**
- **Apesar dos esforços, uma grande parcela da população brasileira não está apta, para se comunicar utilizando LIBRAS, como foi apresentado por Peres et al. (2006). O que gera a dependabilidade de intérpretes para alunos surdos, por exemplo.**

- A inclusão de pessoas com deficiência auditiva e oral ainda é pequena no âmbito global e como foi descrito pela World Health Organization (2023);
- Para o reconhecimento e a tradução automática, grandes desafios de apresentam, como a disponibilidade da bases e o rastreamento de características;
- Dentre as ferramentas que viabilizam o diálogo utilizando LIBRAS, destaca-se o *software* do governo brasileiro, V-LIBRAS.

- Como descrito por Silva et al. (2022), ele utiliza a tradução da língua portuguesa para LIBRAS, por meio do uso de avatares virtuais.
 - Ele apresenta limitações como, a movimentação não fluida dos avatares.



- **Objetivo**

- Este projeto visa desenvolver um sistema para animar avatares virtuais 3D com alta fidelidade, usando sequências de vídeos de sinais da LIBRAS. O foco está na captura de movimentos da parte superior do corpo, incluindo mãos e braços, excluindo a análise de *landmarks* faciais. A animação dos avatares poderá futuramente integrar um sistema de Realidade Virtual ou Aumentada, com potencial para tradução da LIBRAS para o português.

- **Objetivos Específicos**

- Construir um sistema capaz de redirecionar os movimentos de sinais de LIBRAS para avatares humanóides 3D a partir de sequências de vídeos;
- Identificar *datasets* de LIBRAS mais apropriados para o projeto;
- Coletar os vídeos de sinais isolados;
- Realizar experimentos para validação do sistema implementado, a partir de um conjunto selecionado de sinais dinâmicos, obtidos do *dataset* de sinais de LIBRAS.

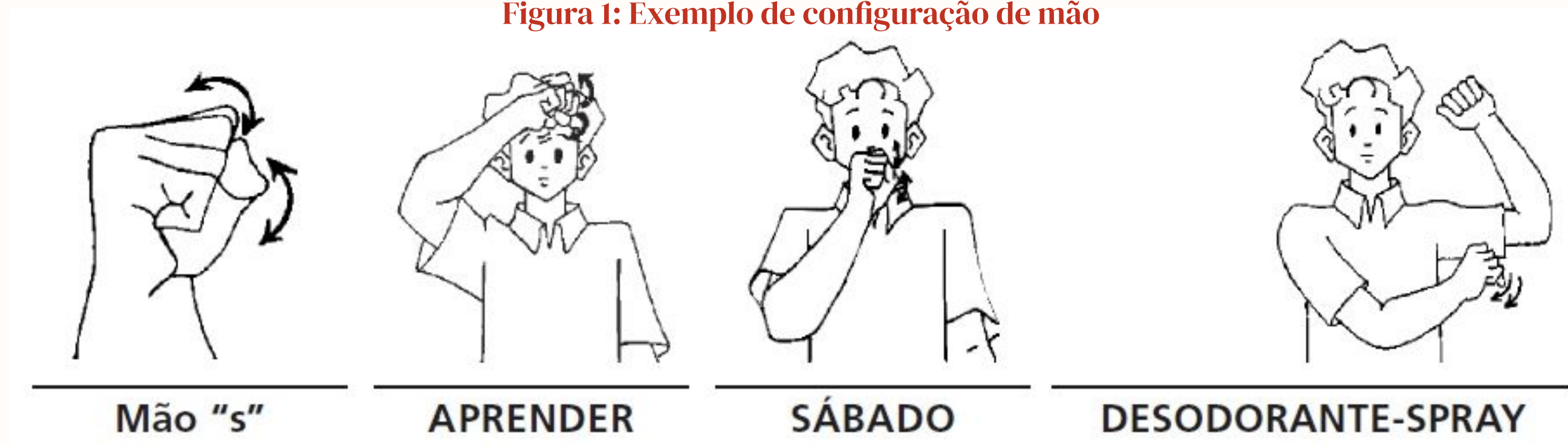
2. Fundamentação Teórica

- Fundamentos da Língua Brasileira de Sinais

- Existem, segundo Capovilla et al. (2017) e Felipe e Monteiro (2007), o total de 5 parâmetros normativos da LIBRAS que são utilizados para a formação de seus sinais.

- Configuração da mão:

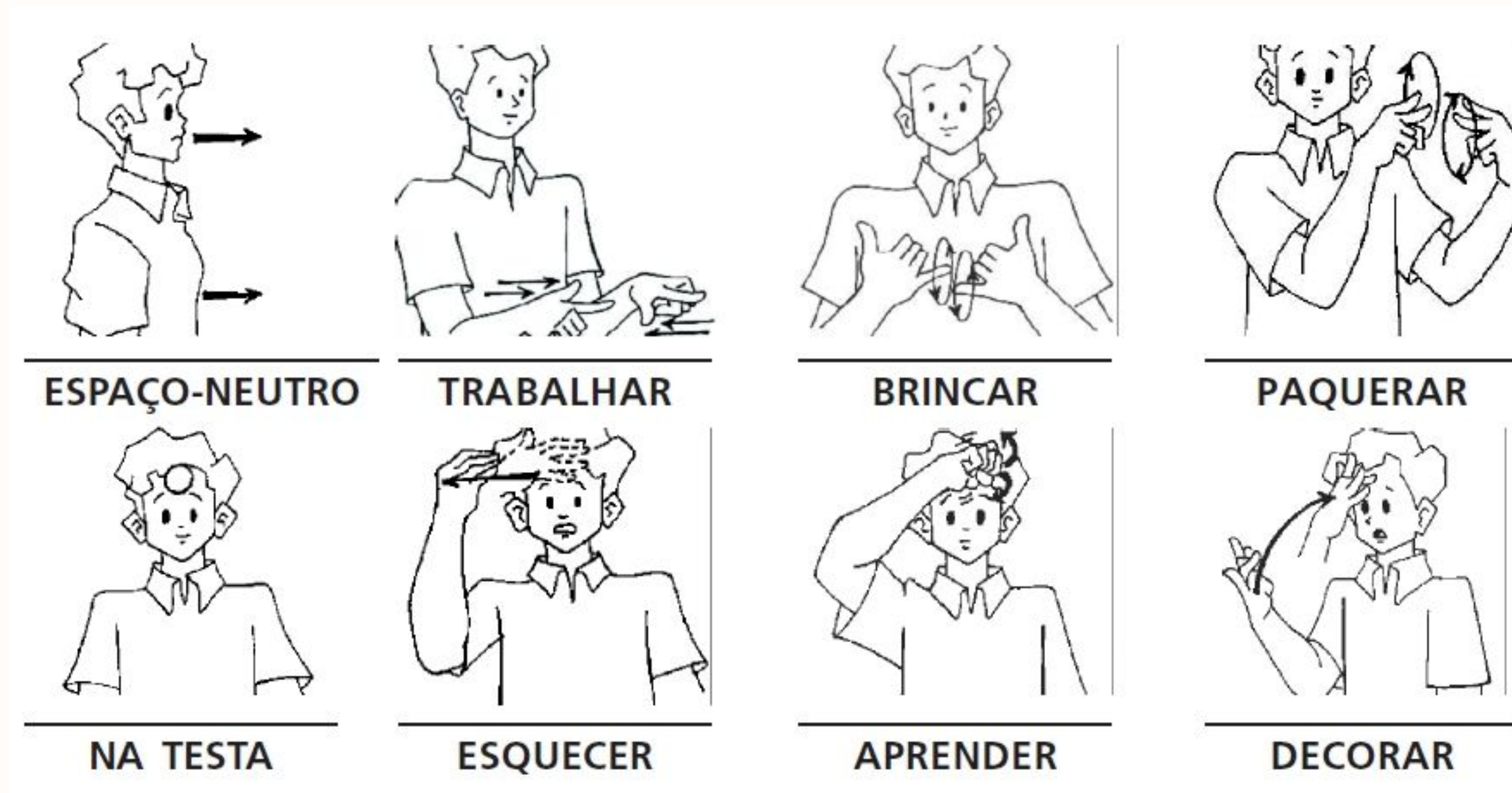
Figura 1: Exemplo de configuração de mão



Fonte: Felipe e Monteiro (2007)

- Fundamentos da Língua Brasileira de Sinais
 - Ponto ou local de articulação:

Figura 2: Exemplo de local de articulação



Fonte: Felipe e Monteiro (2007)

● Fundamentos da Língua Brasileira de Sinais

■ Movimento:

Figura 3: Exemplo de movimento

Têm movimento



RIR



CHORAR



CONHECER

Não têm movimento



AJOELHAR



EM-PÉ

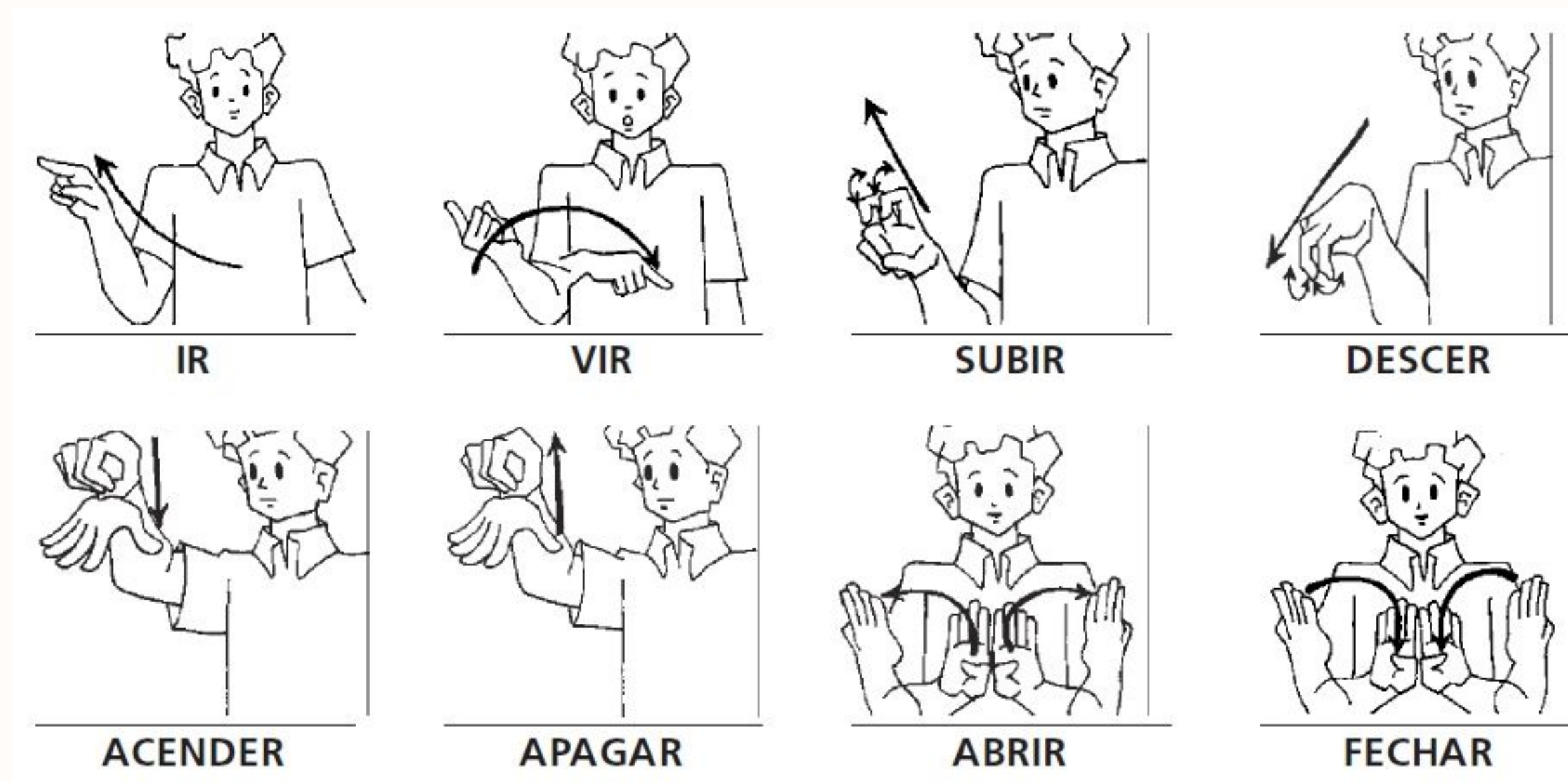


SENTAR

Fonte: Felipe e Monteiro (2007)

- Fundamentos da Língua Brasileira de Sinais
 - Orientação/Direcionalidade:

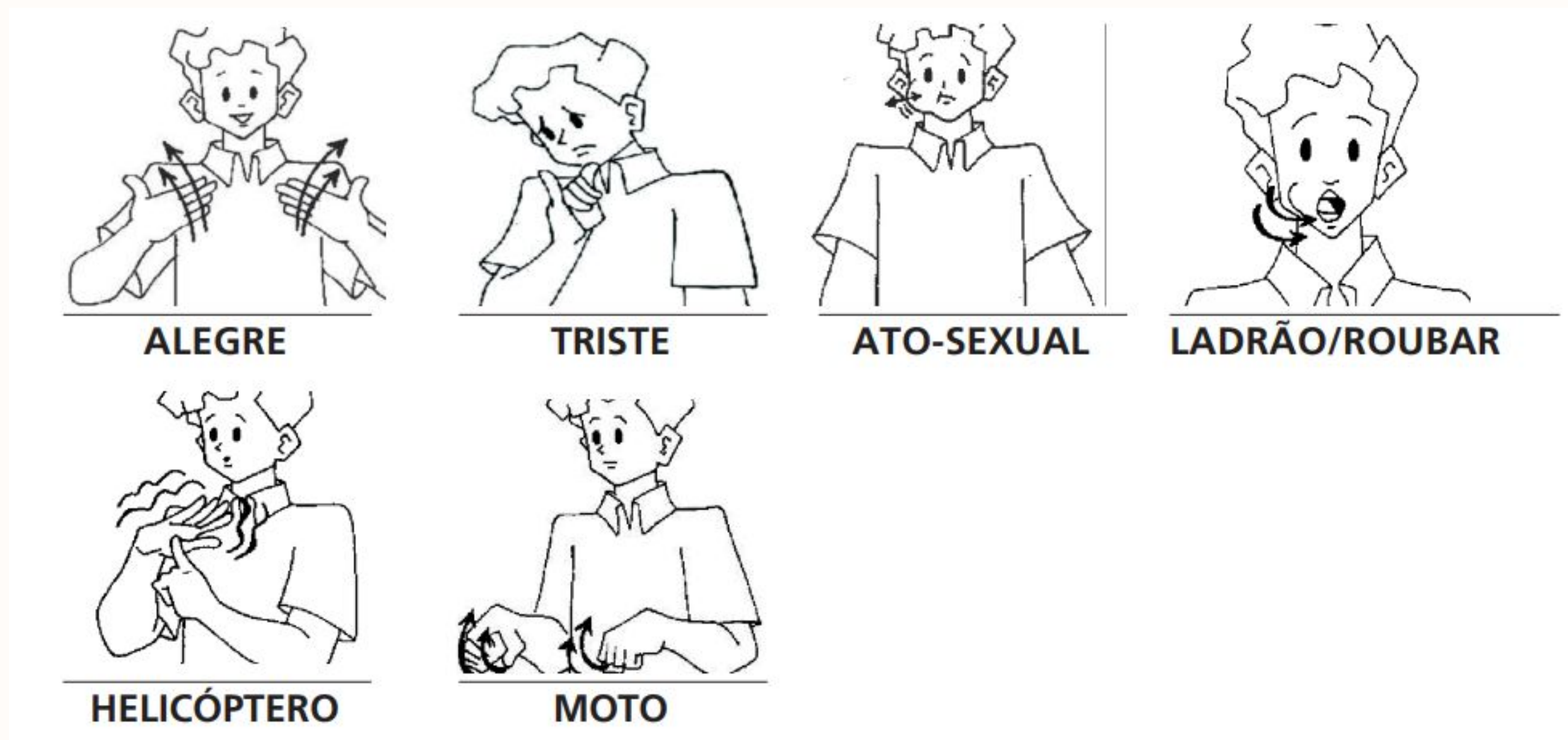
Figura 4: Exemplo de orientação



Fonte: Felipe e Monteiro (2007)

- Fundamentos da Língua Brasileira de Sinais
 - Expressão facial/corporal:

Figura 5: Exemplo de expressão facial/corporal



Fonte: Felipe e Monteiro (2007)

- ***Datasets* de sinais isolados de LIBRAS**
 - Apesar de escassas, principalmente quando comparado com a ASL ou CSL, algumas sólidas bases foram criadas de LIBRAS. As bases consultadas foram:
 - LIBRAS-HC-RGBDS, criada por Porfirio et al. (2013), focou na construção de de uma base centrada nas configurações de mãos.
 - CEFET LIBRAS, apresentado por Gameiro et al. (2020). Essa base possui 547 sequências de vídeos.

- *Datasets* de sinais isolados de LIBRAS
 - LIBRAS-UFOP, de Cerna et al. (2021), em que foram catalogados 56 sinais, divididos pelas semelhanças fonológicas entre si.
 - V-LIBRASIL, publicada por Rodrigues (2021), representa uma grande base de sinais, 1364 palavras.
 - MINDS-Libras, publicada por Rezende (2021), ela possui 20 sinais de LIBRAS gravadas diferentes vezes.
 - Além dessas bases, Sarmiento e Ponti (2023) apresentaram um *dataset* que combina diversas bases de dados, incluindo algumas citadas anteriormente.

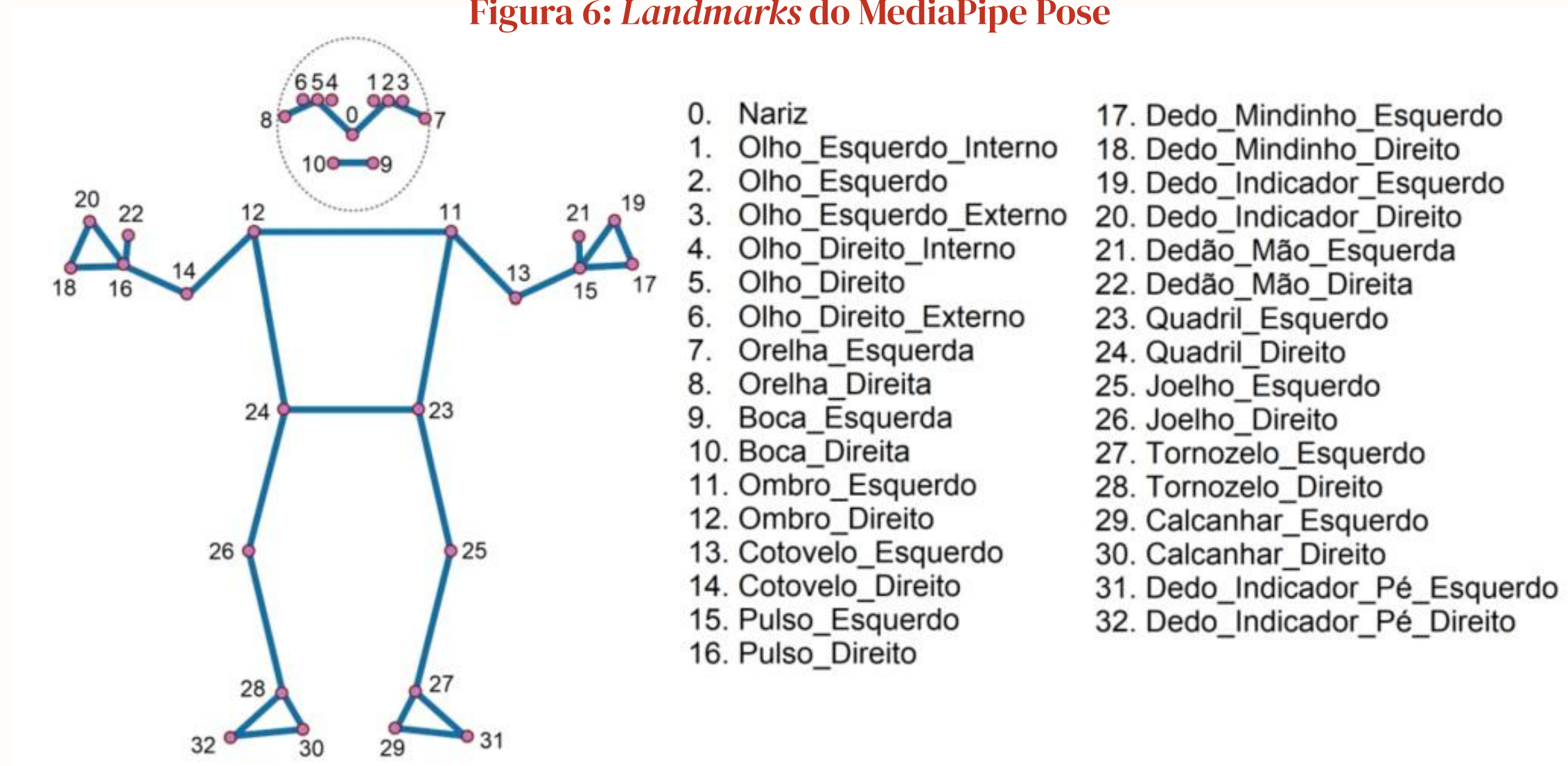
- **Biblioteca MediaPipe**

- É uma solução de bibliotecas de código aberto que auxiliam na aplicação rápida de inteligência artificial e aprendizado de máquina.
- Muito utilizada para o rastreamento de atividades físicas e interpretação de diferentes línguas de sinais.
- Seu funcionamento se dá por uma rede neural convolucional disponibilizado na biblioteca do BlazePose, descrita por Bazarevsky et al. (2020).

- Biblioteca MediaPipe

- Ela consegue extrair pontos chaves, baseados em cada solução. Como o Pose, por exemplo.

Figura 6: Landmarks do MediaPipe Pose



Fonte: Adaptada de Google AI Edge (2023)

- **Realidade Virtual e Avatares 3D**

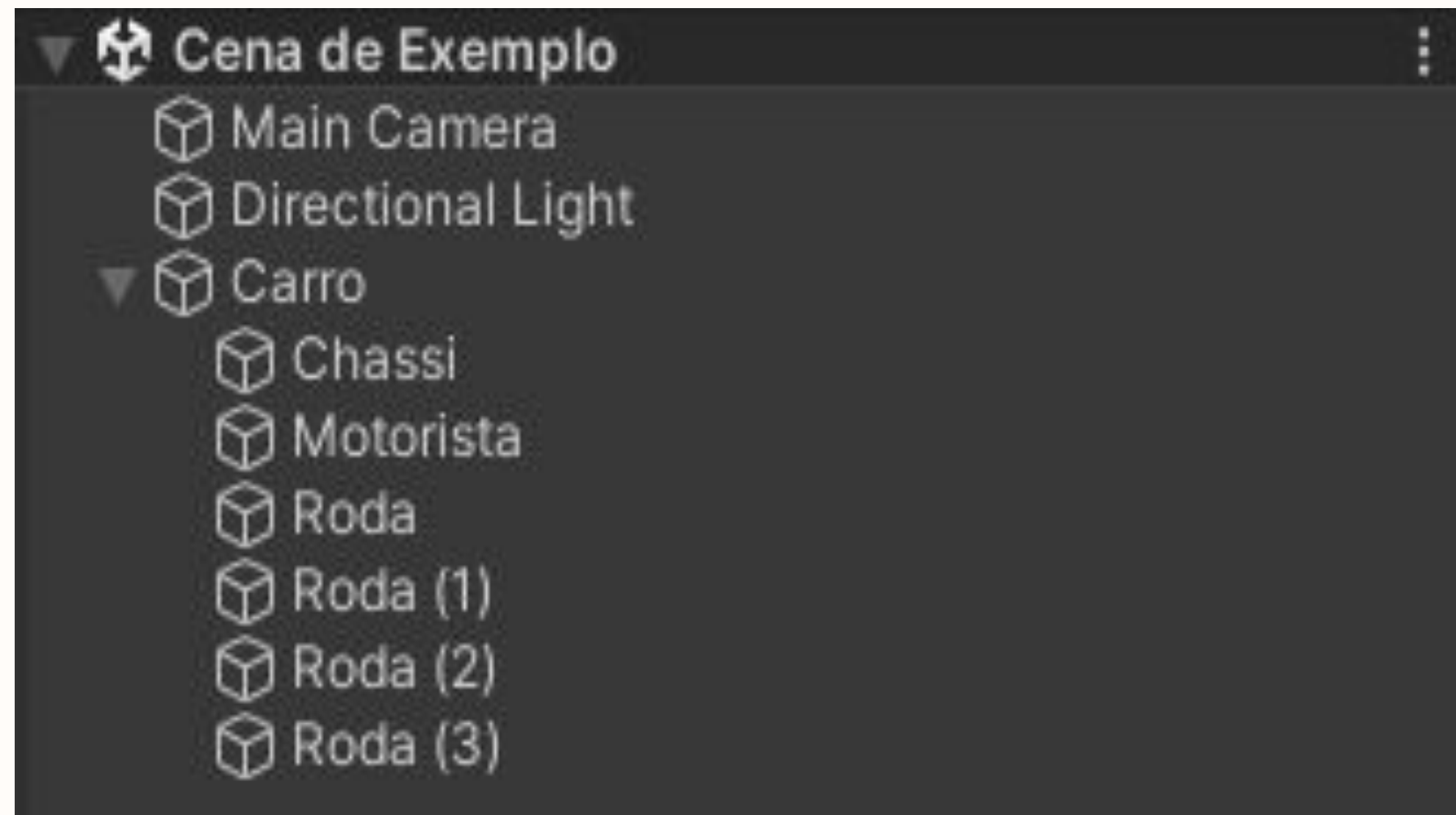
- Para a animação dos sinais de LIBRAS, avatares 3D foram utilizados dentro do mundo virtual.
- O mundo virtual é um ambiente em que o mundo real não existe, tudo que acontece e existe ali é gerado por computador, um ambiente controlado.
- Os avatares 3D são representações dos indivíduos, dentro desse mundo virtual.

- **Motor de Jogos Unity3D**

- O Unity é um motor de jogos que permite o desenvolvimento de ambientes em um cenário virtual tanto 2D quanto 3D (Unity Technologies, 2022);
- Ele permite a integração e adição de criações de outros desenvolvedores, o que permitiu o desenvolvimento deste trabalho;
- Um de seus pontos mais importantes é a organização hierárquica do ambiente desenvolvido.

- **Motor de Jogos Unity3D**

Figura 7: Exemplo de hierarquia no Unity3D



Fonte: Elaborada pelo autor.

- **Motor de Jogos Unity3D**

- Além disso, ele permite modificar diferentes propriedades de cada um dos objetos que estão na cena virtual a partir da componentes que podem ser aplicados a cada objeto;
- Esses componentes podem ser físicos, *scripts*, materiais, entre outros.

- **Trabalhos Correlatos**

- **Dentre o estado da arte associado ao tema, temos os seguintes trabalhos:**
 - **Apresentado por Aberman et al. (2020), ele descreve a criação de uma estrutura de uma rede neural para redirecionamento de movimentos capturados em diferentes esqueletos. Chamada Skeleton-Aware Network, essa rede neural permite o mapeamento entre esqueletos de diferentes estruturas, desde que topologicamente equivalentes.**

- **Trabalhos Correlatos**

- **Song et al. (2023) apresentou sobre um novo método para o redirecionamento de movimento para avatares com diferentes esqueletos e topologias. Nele, um sistema desenvolveu um esqueleto *proxy* unificado que serve como uma estrutura para mapear o esqueleto de entrada e finalizar com os avatares animados.**

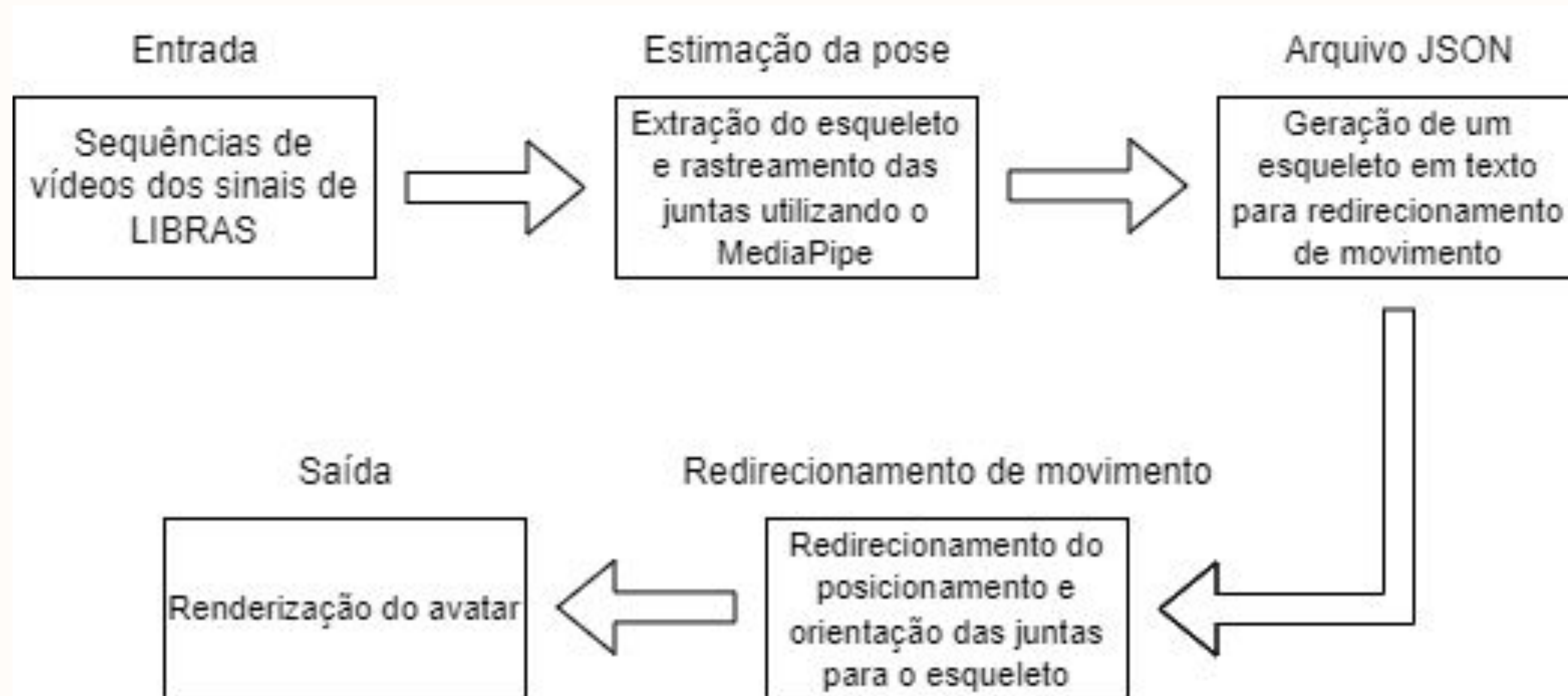
- **Trabalhos Correlatos**

- Por fim, Poulios et al. (2024) apresentou recentemente um método para transferir movimentos de humanos para avatares 3D usando apenas uma câmera RGB. Ele primeiramente estima a posição dos *landmarks* usando o MediaPipe, corrige as imprecisões usando duas redes neurais profundas e por fim transfere o movimento para o avatar no Unity3D.

3. Metodologia

- Método Proposto

Figura 8: Pipeline do sistema proposto



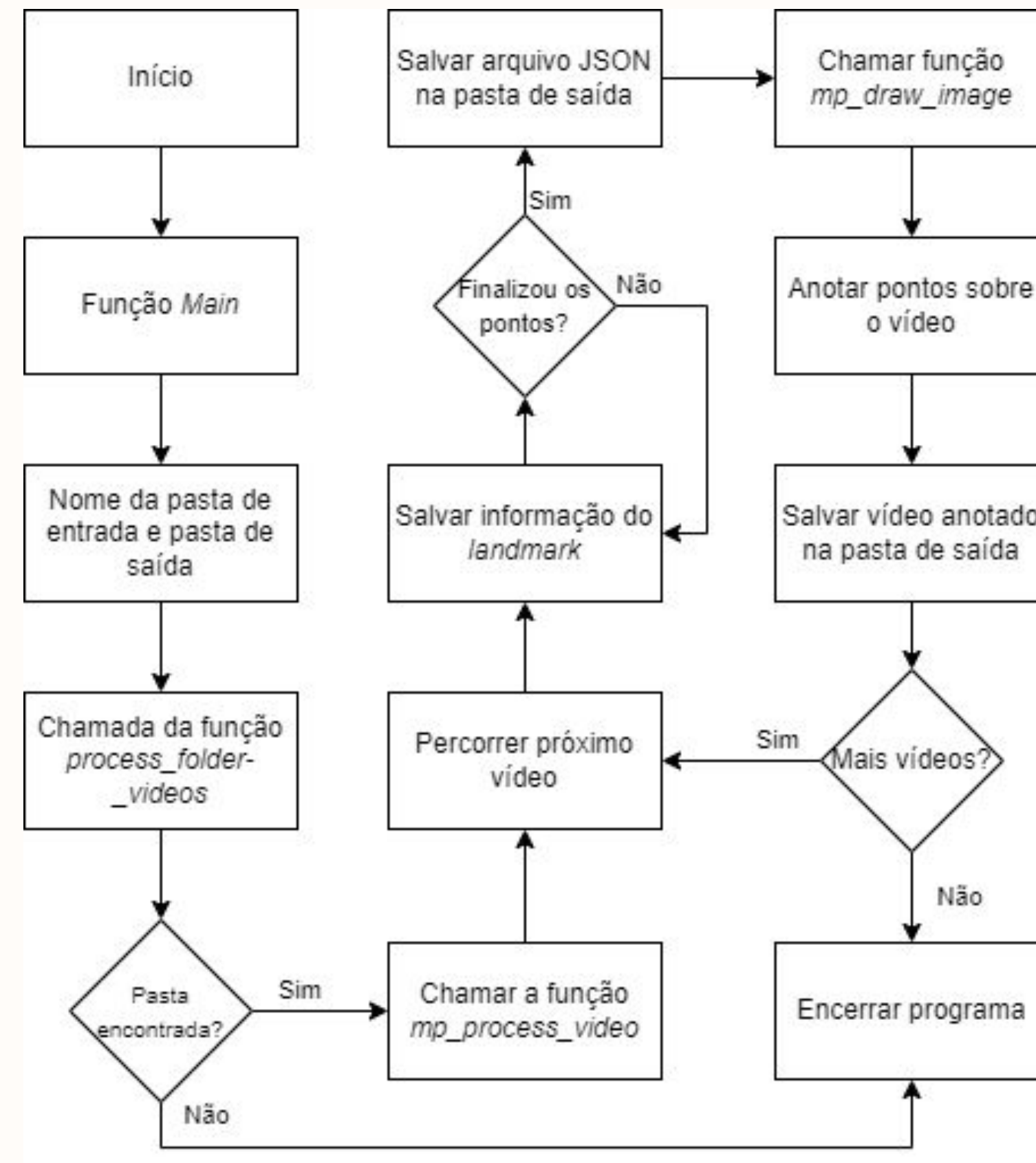
Fonte: Elaborada pelo autor.

- Implementação do Protótipo

- Para se iniciar o desenvolvimento, foi então escolhida o *dataset* V-LIBRASIL, e em seguida, criado o primeiro programa para este projeto;
- O programa, desenvolvido em linguagem Python deve selecionar uma pasta com vídeos de sinais isolados e processar todos utilizando o modelo escolhido do MediaPipe, resultando em um arquivo JSON com a posição dos *landmarks* e o vídeo resultante do processo.

- Implementação do Protótipo
 - Ao lado, o fluxograma com o funcionamento básico do programa.

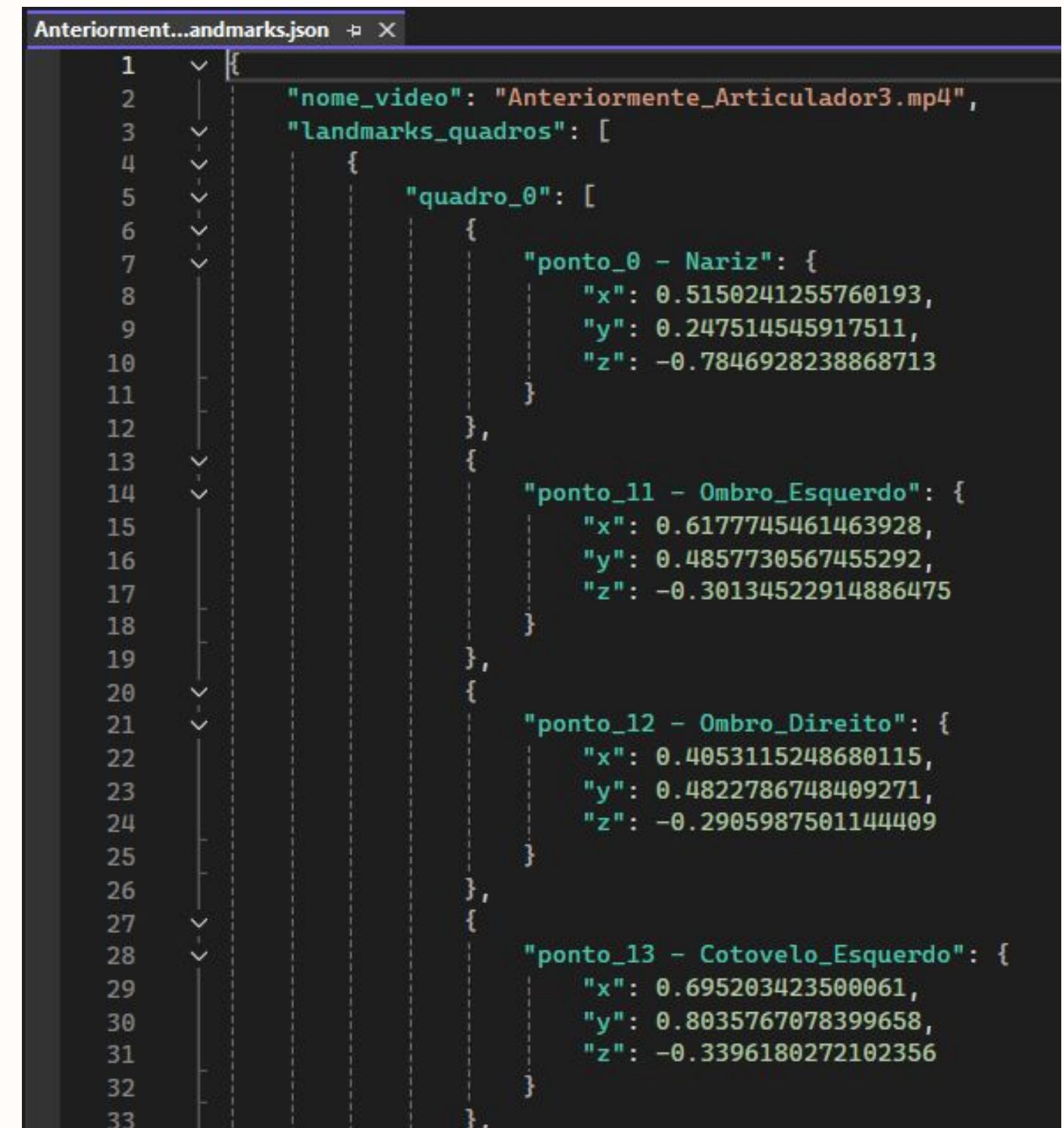
Figura 9: Fluxograma do primeiro programa desenvolvido



Fonte: Elaborada pelo autor.

Figura 10: Estrutura do arquivo JSON gerado

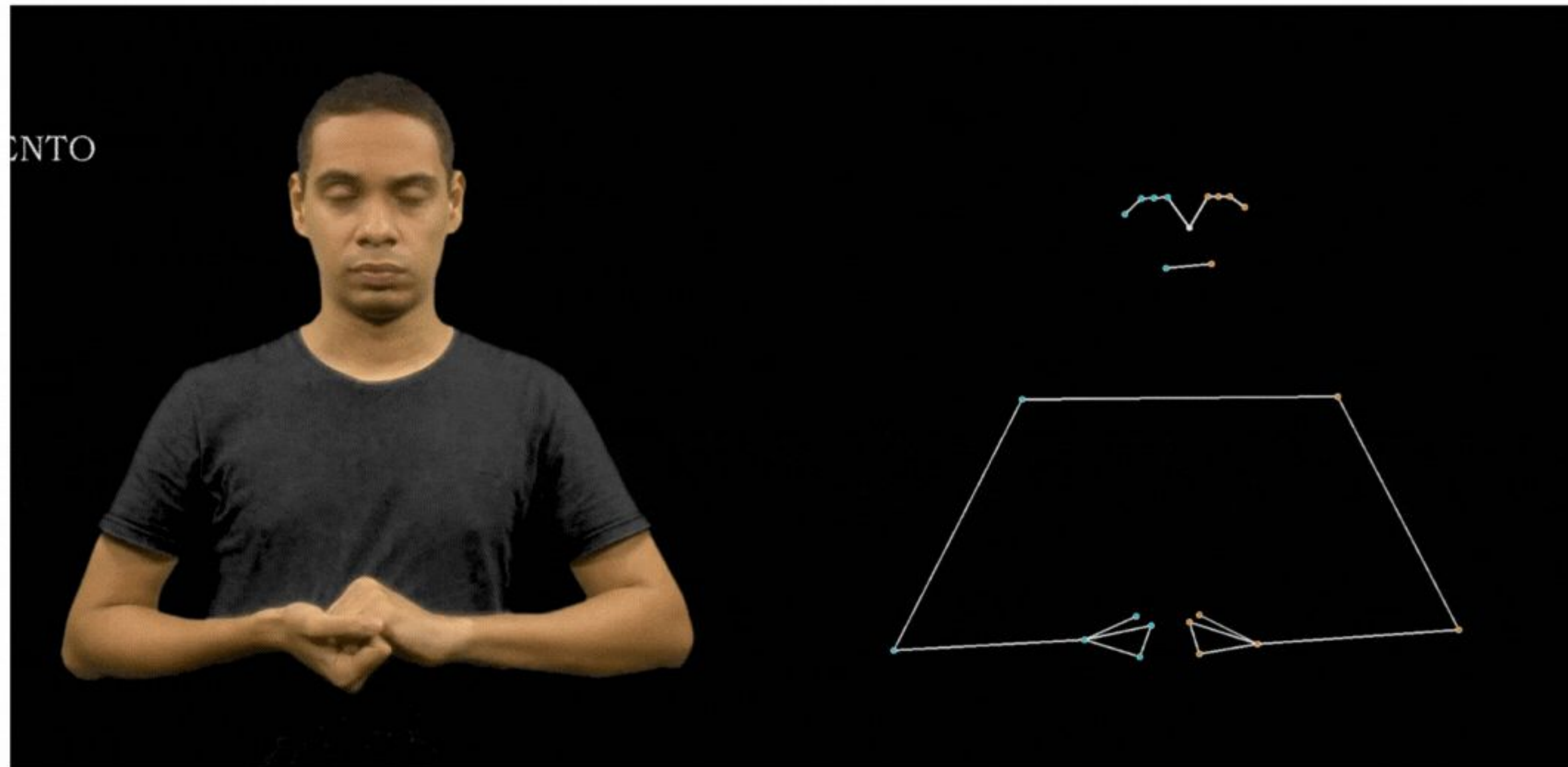
- Implementação do Protótipo
 - Após o processamento, o arquivo JSON sairá com uma estrutura padronizada pronta para ser inserida e processada pelo segundo código desenvolvido.



```
Anteriorment...andmarks.json
1  {
2    "nome_video": "Anteriormente_Articulador3.mp4",
3    "landmarks_quadros": [
4      {
5        "quadro_0": [
6          {
7            "ponto_0 - Nariz": {
8              "x": 0.5150241255760193,
9              "y": 0.247514545917511,
10             "z": -0.7846928238868713
11           }
12         },
13         {
14           "ponto_11 - Ombro_Esquerdo": {
15             "x": 0.6177745461463928,
16             "y": 0.4857730567455292,
17             "z": -0.30134522914886475
18           }
19         },
20         {
21           "ponto_12 - Ombro_Direito": {
22             "x": 0.4053115248680115,
23             "y": 0.4822786748409271,
24             "z": -0.2905987501144409
25           }
26         },
27         {
28           "ponto_13 - Cotovelo_Esquerdo": {
29             "x": 0.695203423500061,
30             "y": 0.8035767078399658,
31             "z": -0.3396180272102356
32           }
33         }
34       ]
35     }
36   ]
37 }
```

Fonte: Elaborada pelo autor.

Figura 11: Exemplo de vídeo anotado e vídeo original



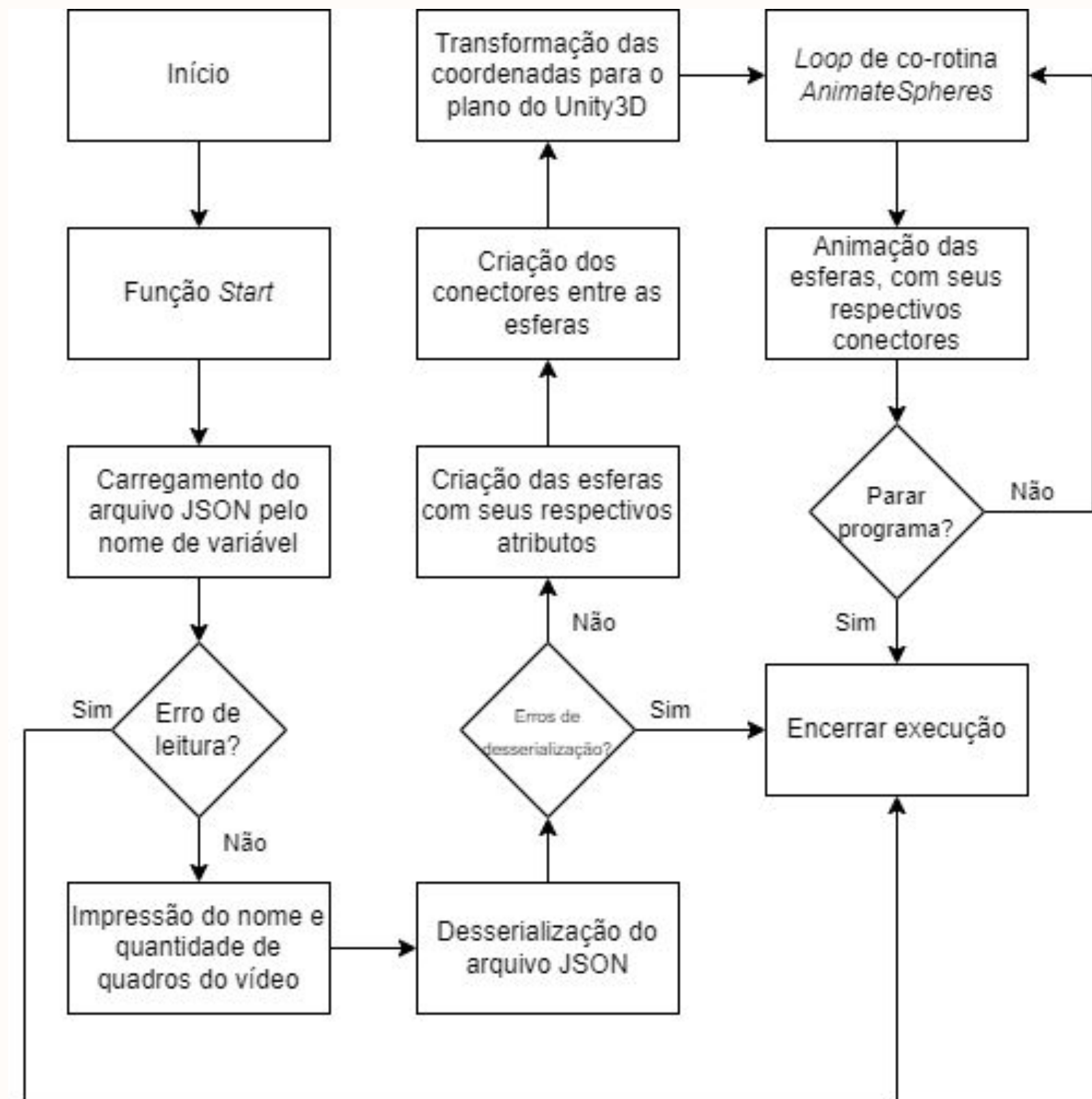
Fonte: Elaborado pelo autor.

- **Implementação do Protótipo**

- Já para o segundo programa é responsável pela renderização do avatar e pela desserialização do arquivo JSON, já dentro do ambiente do Unity3D.
- Ele consiste em um programa que recebe o arquivo JSON, faz a desserialização, e redireciona as coordenadas dos pontos para esferas dentro do ambiente 3D. Utilizando ainda os dados do arquivo JSON, ele movimento essas esferas, e suas juntas, para animar o sinal desejado.

- Implementação do Protótipo
 - Ao lado, o fluxograma com o funcionamento básico do programa.

Figura 12: Fluxograma do segundo programa desenvolvido

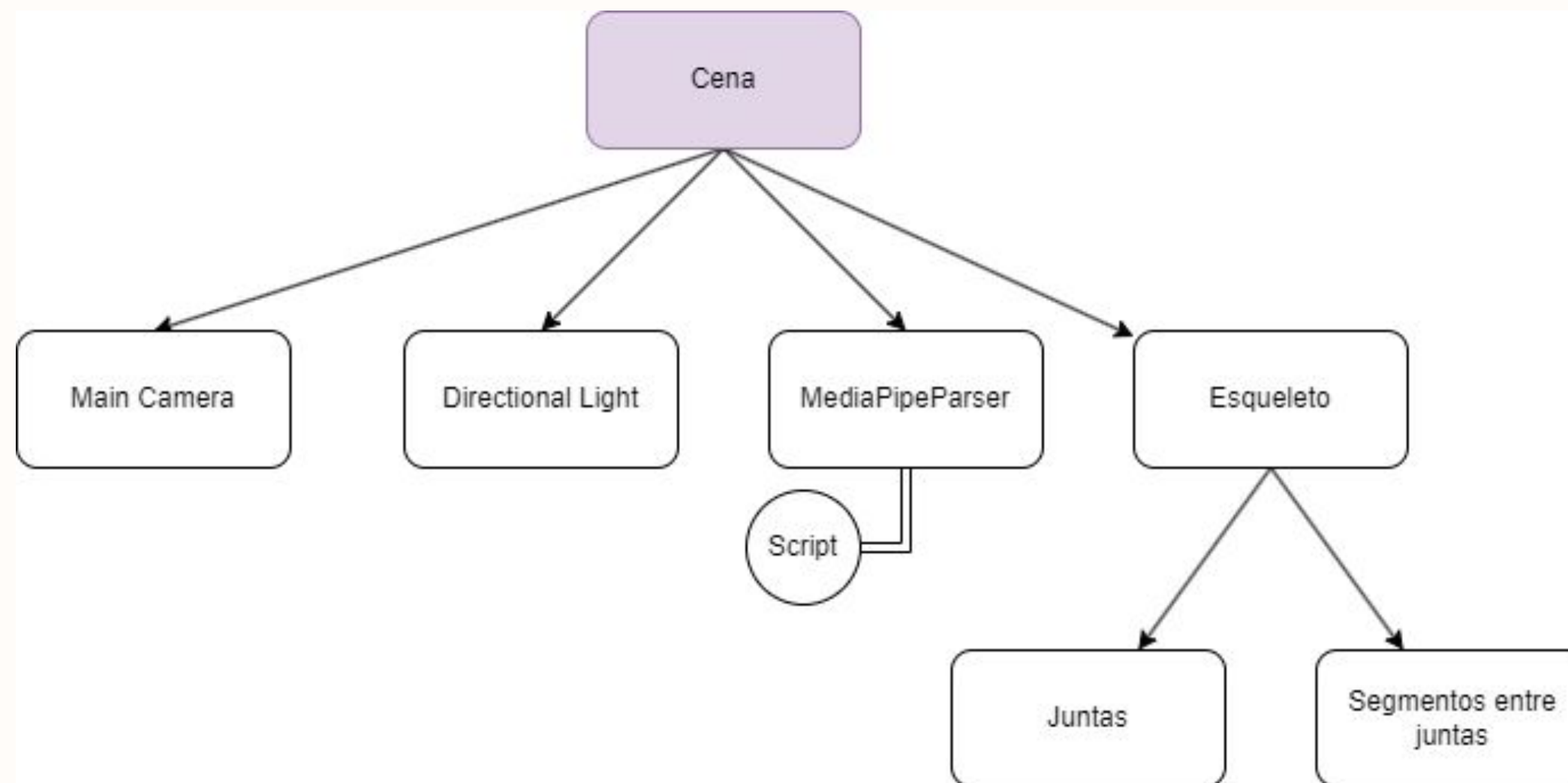


Fonte: Elaborada pelo autor.

- Implementação do Protótipo

- Abaixo, está o grafo de cena do ambiente desenvolvido.

Figura 13: Grafo de cena do ambiente desenvolvido



Fonte: Elaborada pelo autor.

4. Experimentos

- Interface

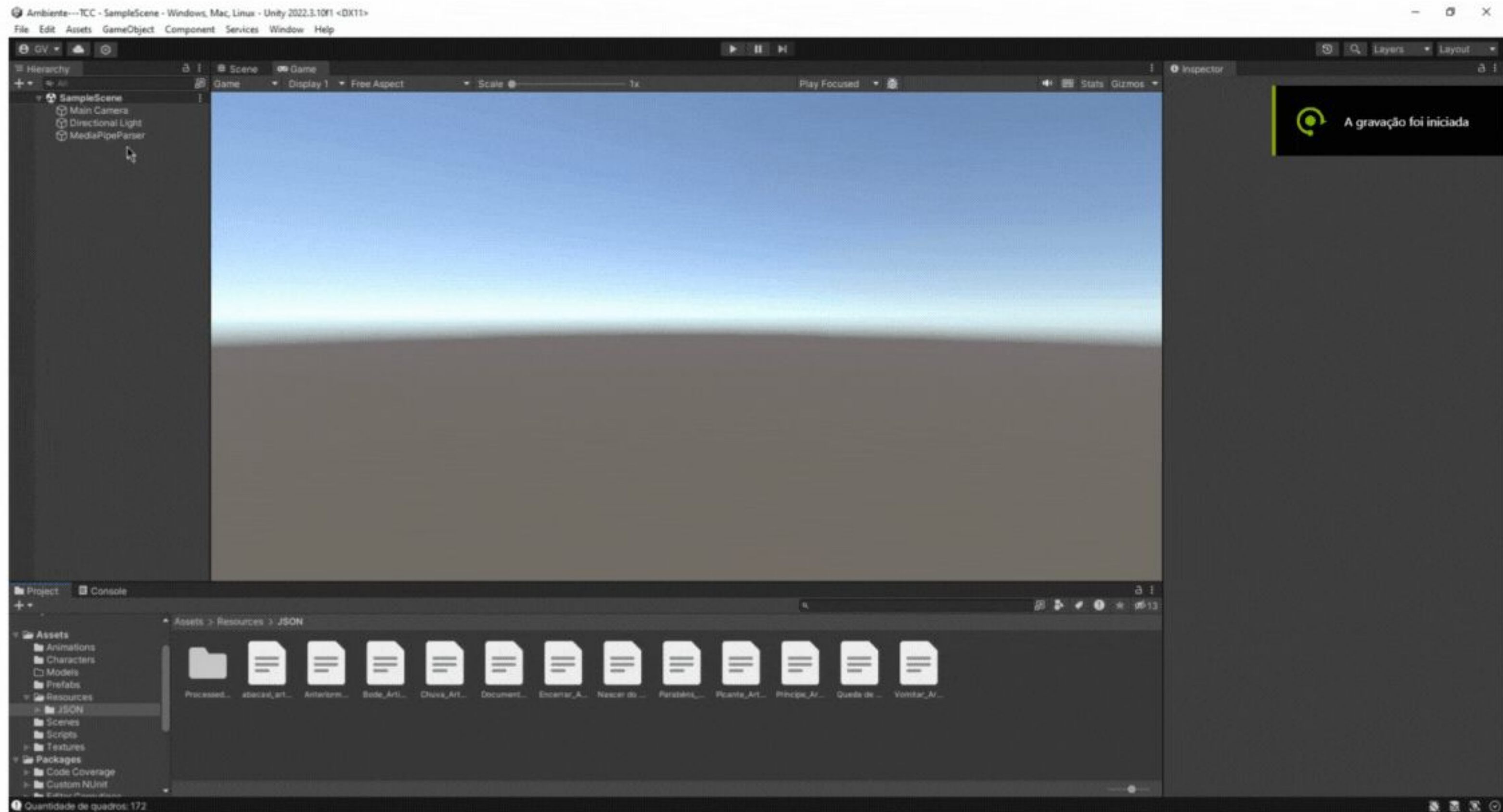
- Para os experimentos, foi utilizada a interface gráfica do Unity3D. E o componente de *script* permite que diferentes variáveis pudessem ser customizadas, incluindo o sinal desejado para a animação.

Figura 14: Interface das variáveis no Unity3D



Fonte: Elaborada pelo autor.

Figura 15: Exemplo do programa em execução



Fonte: Elaborada pelo autor.

- **Experimentos**

- **Foram realizados 12 experimentos, com 12 sinais diferentes de LIBRAS.**

- **Abacaxi;**
- **Anteriormente;**
- **Bode;**
- **Chuva;**
- **Documento;**
- **Encerrar;**
- **Nascer do Sol;**
- **Parabéns;**

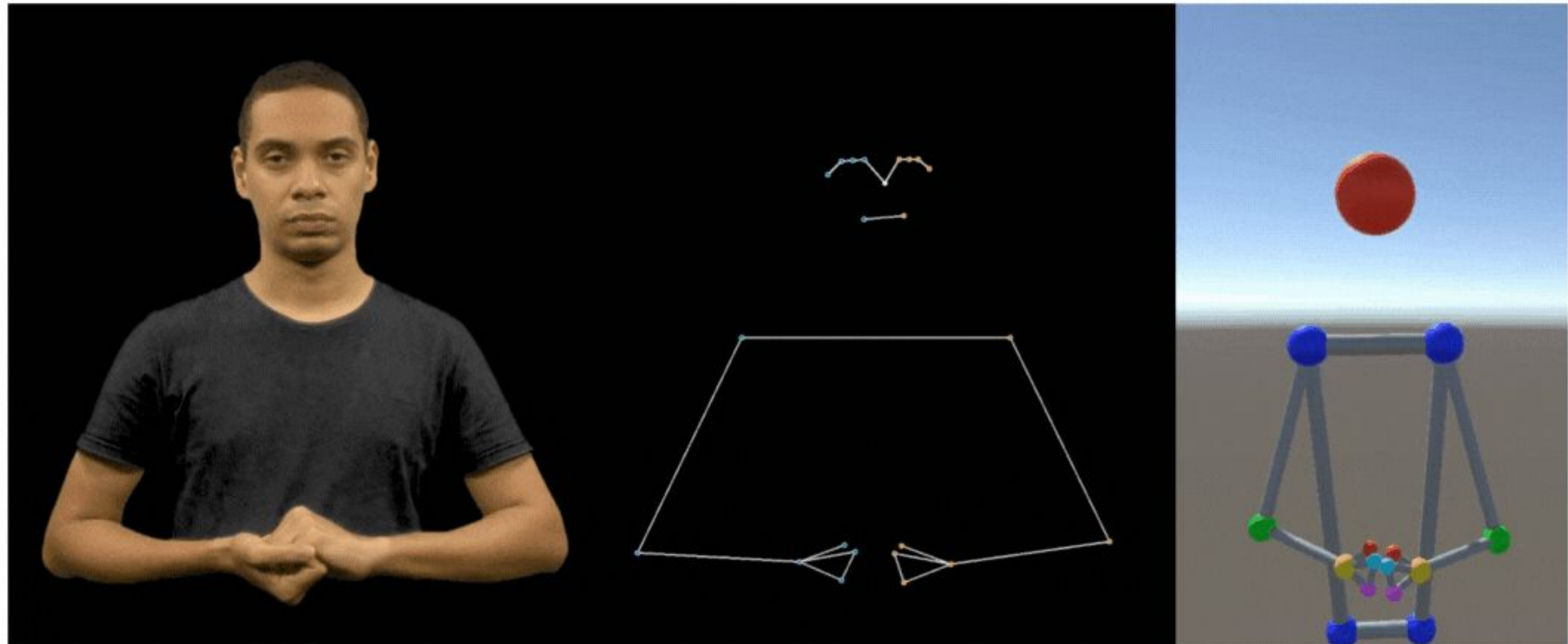
- **Picante;**
- **Príncipe;**
- **Queda de Energia;**
- **Vomitar.**

Figura 16: Exemplo do programa em execução com dois exemplos



Fonte: Elaborada pelo autor.

Figura 17: Vídeo original, vídeo anotado e esqueleto do avatar animado



Fonte: Elaborada pelo autor.

5. Conclusões

- **Conclusões:**
 - Foi desenvolvido um ambiente virtual capaz de animar sinais de LIBRAS, com foco nos movimentos de mãos e braços;
 - Utilizando o ambiente virtual desenvolvido, foram realizados testes com 12 sinais de LIBRAS, mas o método desenvolvido pode ser utilizado para qualquer base de dados ou sinais isolados de LIBRAS;
 - Além disso, um algoritmo capaz de fazer o processamento de qualquer *dataset* também foi desenvolvido. Permitindo o uso do ambiente criado no Unity.

- **Trabalhos Futuros:**
 - Aplicação do SLERP para suavizar transições nas animações;
 - Desenvolvimento de uma topologia aprimorada para o esqueleto do avatar, conforme sugerido por Song et al. (2023);
 - Testar modelos avançados do MediaPipe, como o Holistic, e testar diferentes *datasets* de LIBRAS;
 - Aplicar uma rede neural para a estimação de dados de profundidade com maior precisão, como apresentado por Poulios et al. (2024);
 - Realizar uma avaliação de inferência com métricas mais objetivas.

Referências

- BRASIL. Lei nº 10.436, de 24 de abril de 2002. Dispõe sobre a Língua Brasileira de Sinais - Libras e dá outras providências. Brasília, DF, 2002. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/2002/l10436.htm. Acesso em: 16 mar. 2024.
- CAPOVILLA, F. C.; RAPHAEL, W. D.; MARTINS, A. C.; TEMOTEO, J. G. Dicionário da Língua de Sinais do Brasil: A Libras em suas Mãos. 1. ed. [S.l.]: EDUSP, 2017. ISBN 978-85-314-1645-3
- CRISTIANO, A. "O que é Libras?". 2017. Disponível em: <https://www.libras.com.br/o-que-e-libras>. Acesso em: 16 mar. 2024.
- PERES, S. M.; FLORES, F. C.; VERONEZ, D.; OLGUIN, C. J. M. Libras signals recognition: a study with learning vector quantization and bit signature. In: 2006 Ninth Brazilian Symposium on Neural Networks (SBRN'06). [s.n.], 2006. p. 119–124. Disponível em: <https://ieeexplore.ieee.org/document/4026821>. Acesso em: 23 mar. 2024.

- **World Health Organization. Deafness and hearing loss. 2023. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>. Acesso em: 16 mar. 2024.**
- **SILVA, A. L. d. C.; Sá, T. M. d.; DINIZ, R. S.; FERREIRA, S. B. L.; SIQUEIRA, S. W. M.; BOURGUIGNON, S. C. Prescriptive and Semantic Analysis of an Automatic Sign Language Translation: Cases on VLibras Avatar Translation Using Video Interviews and Textual Interactions With a Chatbot. *Interacting with Computers*, v. 35, n. 2, p. 231–246, 10 2022. ISSN 1873-7951. Disponível em: <https://doi.org/10.1093/iwc/iwac020>. Acesso em: 17 mar. 2024.**
- **FELIPE, T. A.; MONTEIRO, M. Libras em Contexto: Curso Básico : Livro do Professor. 6. ed. [S.l.]: WalPrint, 2007.**
- **PORFIRIO, A. J.; WIGGERS, K. L.; OLIVEIRA, L. E.; WEINGAERTNER, D. Libras sign language hand configuration recognition based on 3d meshes. In: 2013 IEEE International Conference on Systems, Man, and Cybernetics. [s.n.], 2013. p. 1588–1593. Disponível em: <https://ieeexplore.ieee.org/document/6722027>. Acesso em: 23 out. 2024.**

- GAMEIRO, P. V.; PASSOS, W. L.; ARAUJO, G. M.; LIMA, A. A. de; GOIS, J. N.; CORBO, A. R. A brazilian sign language video database for automatic recognition. In: 2020 Latin American Robotics Symposium (LARS), 2020 Brazilian Symposium on Robotics (SBR) and 2020 Workshop on Robotics in Education (WRE). [s.n.], 2020. p. 1–6. Disponível em: <https://ieeexplore.ieee.org/document/9307017>. Acesso em: 23 out. 2024.
- CERNA, L. R.; CARDENAS, E. E.; MIRANDA, D. G.; MENOTTI, D.; CAMARA-CHAVEZ, G. A multimodal libras-ufop brazilian sign language dataset of minimal pairs using a microsoft kinect sensor. *Expert Systems with Applications*, v. 167, p. 114179, 2021. ISSN 0957-4174. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0957417420309143>. Acesso em: 23 out. 2024.
- RODRIGUES, A. J. V-LIBRASIL : uma base de dados com sinais na língua brasileira de sinais (Libras). Tese (Dissertação de Mestrado) — Universidade Federal de Pernambuco, ago. 2021. Disponível em: <https://repositorio.ufpe.br/handle/123456789/43491>. Acesso em: 23 mar. 2024.

- REZENDE, T. M. Reconhecimento automático de sinais da Libras : desenvolvimento da base de dados MINDS-Libras e modelos de redes convolucionais. Tese (Tese de Doutorado) — Universidade Federal de Minas Gerais, jul. 2021. Disponível em: <http://hdl.handle.net/1843/39785>. Acesso em: 24 out. 2024.
- SARMENTO, A. H. de A.; PONTI, M. A. A cross-dataset study on the brazilian sign language translation. In: 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). [s.n.], 2023. p. 2808–2812. Disponível em: <https://ieeexplore.ieee.org/document/10350372>. Acesso em: 24 out. 2024.
- BAZAREVSKY, V.; GRISHCHENKO, I.; RAVEENDRAN, K.; ZHU, T.; ZHANG, F.; GRUNDMANN, M. BlazePose: On-device Real-time Body Pose tracking. 2020. Disponível em: <https://arxiv.org/abs/2006.10204>. Acesso em: 23 mar. 2024.
- Google AI Edge. MediaPipe Pose. 2023. Disponível em: <https://github.com/google-ai-edge/mediapipe/blob/master/docs/solutions/pose.md>. Acesso em: 22 out. 2024.

- **Unity Technologies. Unity - Manual: Unity User Manual 2022.3 (LTS). 2022. Disponível em:**
<https://docs.unity3d.com/2022.3/Documentation/Manual/index.html>. Acesso em: 25 out. 2024.
- **ABERMAN, K.; LI, P.; LISCHINSKI, D.; SORKINE-HORNUNG, O.; COHEN-OR, D.; CHEN, B. Skeleton-aware networks for deep motion retargeting. ACM Trans. Graph., Association for Computing Machinery, New York, NY, USA, v. 39, n. 4, ago. 2020. ISSN 0730-0301. Disponível em:**
<https://doi.org/10.1145/3386569.3392462>. Acesso em: 26 out. 2024.
- **POULIOS, I.; PISTOLA, T.; SYMEONIDIS, S.; DIPLARIS, S.; IOANNIDIS, K.; VROCHIDIS, S.; KOMPATSIARIS, I. Enhanced real-time motion transfer to 3d avatars using rgb-based human 3d pose estimation. In: Proceedings of the 2024 ACM International Conference on Interactive Media Experiences Workshops. New York, NY, USA: Association for Computing Machinery, 2024. (IMXw '24), p. 88–99. ISBN 9798400717949. Disponível em:**
<https://doi.org/10.1145/3672406.3672427>. Acesso em: 26 out. 2024

- **SONG, W.; ZHANG, X.; GAO, Y.; LUO, Y.; WANG, H.; WANG, X.; HOU, X. Upsr: a unified proxy skeleton retargeting method for heterogeneous avatar animation. In: 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). [s.n.], 2023. p. 867–868. Disponível em: <https://ieeexplore.ieee.org/document/10108598>. Acesso em: 23 mar. 2024.**

Obrigado pela
atenção