

ASTRA: ASSISTENTE INTELIGENTE PARA NAVEGAÇÃO DE PESSOAS COM DEFICIÊNCIA VISUAL EM AMBIENTES COMPLEXOS

Lucas Montoya Correia¹, Marcelo Xavier Guterres²

¹ Universidade Presbiteriana Mackenzie, São Paulo (SP), Brasil
² Instituto Tecnológico de Aeronáutica, São José dos Campos (SP), Brasil
Contato: lucas.correia.sas@gmail.com

INTRODUÇÃO

Pessoas com deficiência visual enfrentam barreiras de mobilidade em ambientes complexos (aeroportos, shopping centers, vias públicas). Soluções acessíveis ainda têm cobertura funcional limitada. A Astra integra um vestível leve (câmera, microfone e sensores) a um backend em tempo real para oferecer descrição de cenas, leitura de textos, navegação interna por imagens de referência, validação visual externa e alerta de obstáculos por áudio espacial. Este trabalho apresenta um protótipo validado em laboratório e pronto para testes de campo.

OBJETIVOS

- Primário:
- Promover autonomia com experiência de uso fluida por meio de um assistente vestível multimodal em tempo real.
- Secundários:
- Aplicar visão computacional à mobilidade assistida (descrição, leitura, navegação, reconhecimento).
 - Interação natural via voz com STT/TTS no servidor e migração futura para cloud.
 - Garantir modularidade de hardware/software e baixo custo de adoção.
 - Avançar na miniaturização e otimização energética.

METODOLOGIA

Arquitetura do Sistema:

Vestível conectado ao servidor orquestra visão e voz em tempo real (Fig. 1).

- Hardware (vestível): câmera, microfone digital (I²S), IMU, ToF, GPS, bateria Li-Po + power-path; retorno por fone Bluetooth
- Pipeline: STT → LLM (orquestrador) → Módulos de Visão/Navegação → TTS
- Conectividade: hotspot/Wi-Fi; pareamento por voz ou QR Code
- Módulos principais: descrição de cena; OCR/leitura; navegação por imagens-âncora; validação visual externa; alerta de obstáculos com áudio espacial

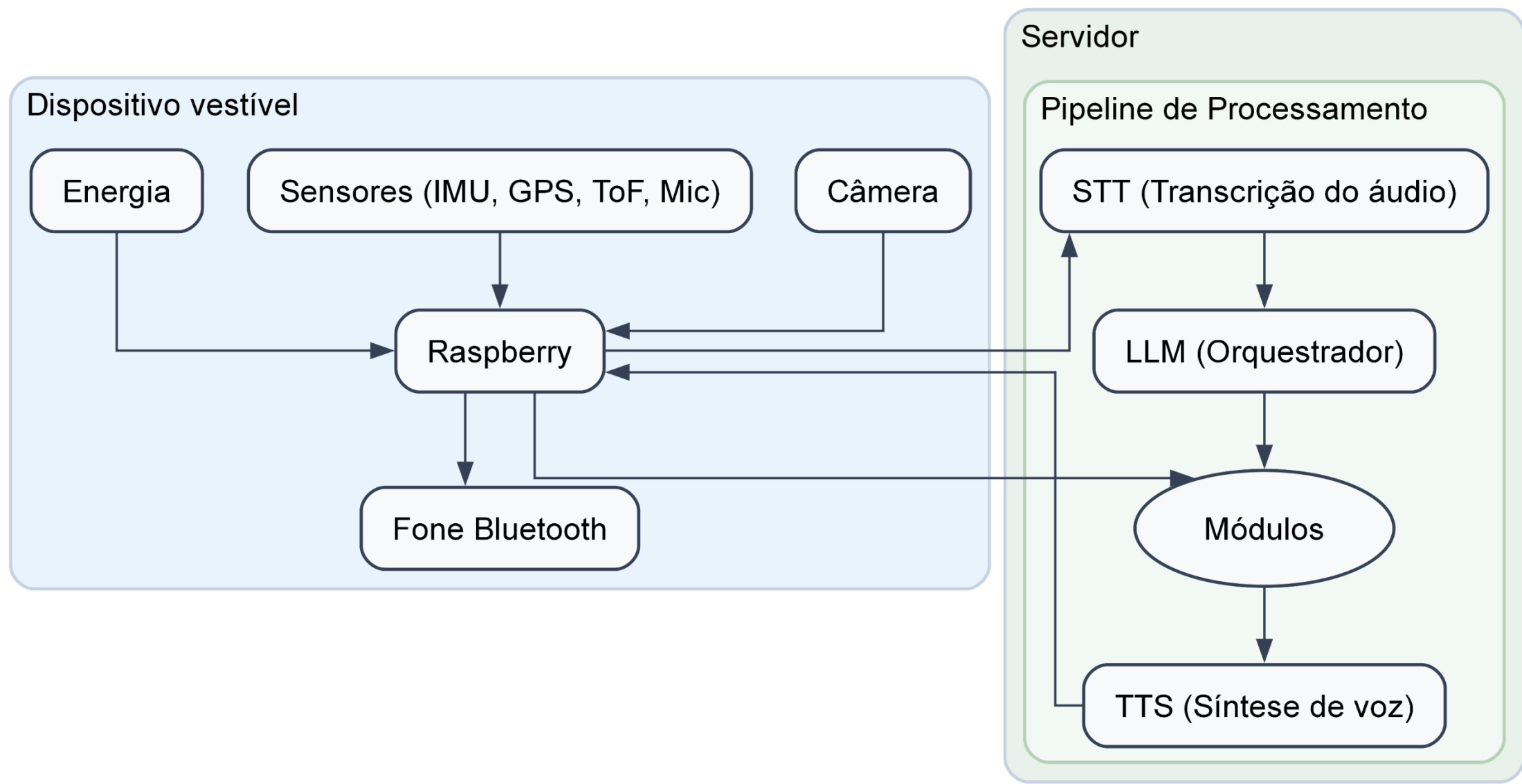


FIGURA 1: Arquitetura do sistema

A Raspberry envia áudio ao servidor (STT) e vídeo+sensores aos módulos de Visão/Navegação; a resposta sintetizada (TTS) retorna ao fone Bluetooth.

Fluxo de Interação:

- Passos do ponto de vista do usuário e do sistema.
- Usuário diz a tarefa (“ir à loja X” / “ler placa” / “procurar objeto/rosto”)
 - STT transcreve e envia ao LLM
 - LLM seleciona módulos adequados (ex.: OCR + navegação interna)
 - Processamento visual executa a tarefa e confirma marcos quando aplicável (validação externa)
 - Sistema gera instruções auditivas com áudio espacial (direção/distância)
 - TTS sintetiza a resposta e envia ao fone Bluetooth

Métricas e Protocolo:

- Avaliamos o protótipo em ambiente controlado, executando rotas/tarefas repetidas com o pipeline completo.
- Latência fala→fala: ~5 s (início da fala → início do TTS).
 - Vídeo (pipeline): 5–10 FPS.
 - Autonomia: ≥ 1,5 h (uso contínuo).
 - Navegação interna: consistente no cenário controlado; testes de campo pendentes para validação externa.
 - Como medimos: cronômetro automatizado de fala→TTS; FPS médio do pipeline; autonomia sob uso realista (visão ativa + voz).

RESULTADOS E DISCUSSÃO

- Operação em tempo real no laboratório com pipeline completo
- Navegação interna consistente em cenário controlado; validação externa integrada, pendente de campo
- Métricas: latência média 5 s; vídeo 5–10 FPS; autonomia ≥ 1,5 h (varia com carga/uso)
- Feedback exploratório: adolescente com deficiência visual destacou navegação interna como recurso de maior interesse (ex.: shoppings)
- Limitações atuais: latência dependente do backend local; miniaturização; consumo sob percepção contínua
- Próximos passos (12 meses): migrar para cloud, app Wi-Fi Direct, upgrade da placa de processamento, PCB flex
- Escalabilidade responsável: modularidade permite expansão para outros públicos (crianças, idosos, diferentes perfis de usuários) e outras aplicações (saúde, robótica social, drones e veículos autônomos)
- Propriedade intelectual: elementos sensíveis protegidos; avaliação de abertura parcial de artefatos para colaboração acadêmica

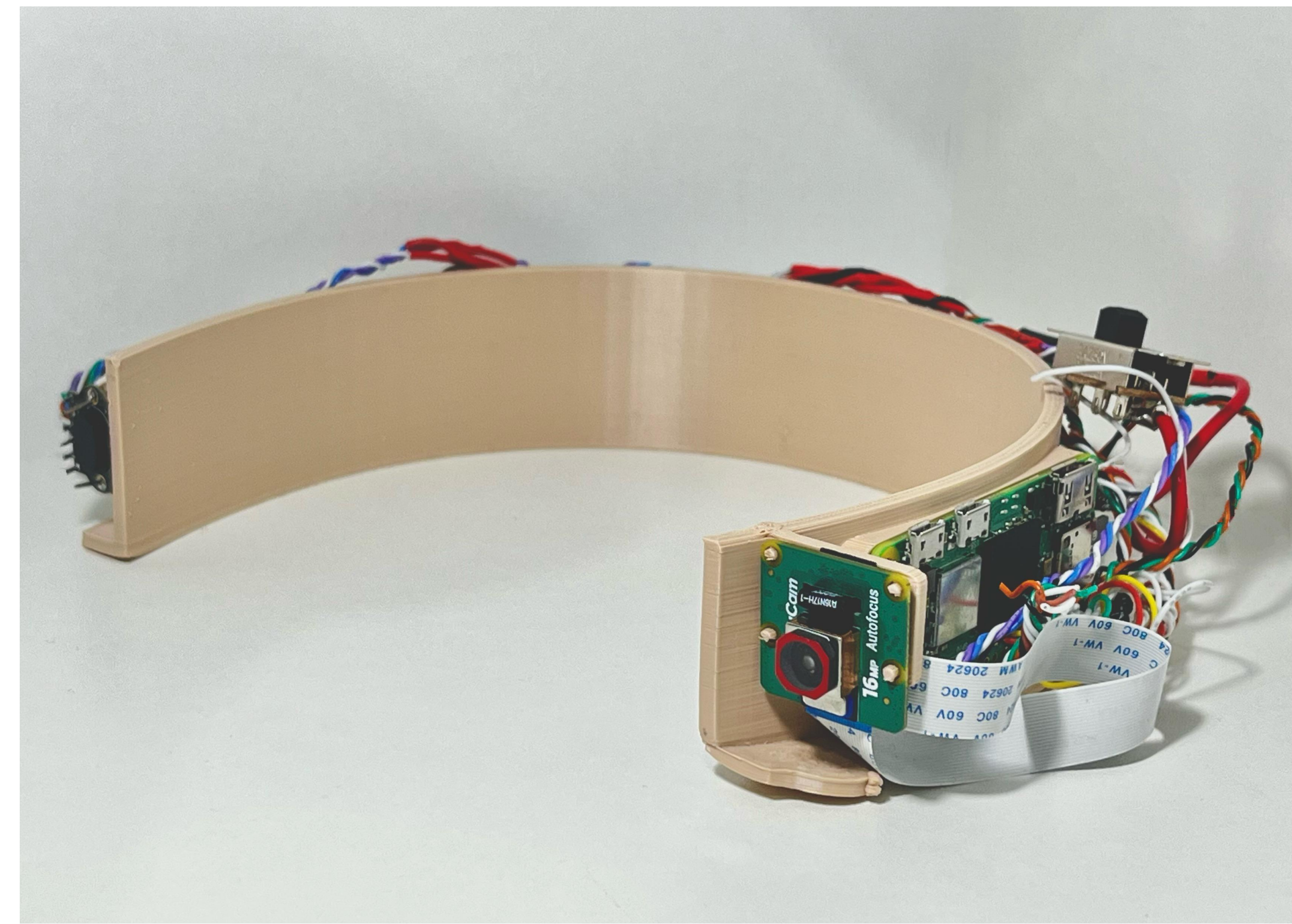


FIGURA 2: Protótipo vestível (v0.2)

Câmera, microfone digital, IMU, GPS, ToF, SD e LiPo + power-path em estrutura 3D.

CONCLUSÕES

A Astra demonstrou viabilidade técnica em laboratório, reunindo navegação interna guiada por marcos visuais, validação externa integrada e interação por voz em tempo real (5 s; 5–10 FPS; ≥1,5 h). A migração do backend para cloud e a miniaturização visam maior previsibilidade, estabilidade e adoção cotidiana. O enfoque em navegação interna e arquitetura modular diferencia a proposta e sustenta o potencial de impacto social.

AGRADECIMENTOS

Agradecemos ao CNPq (Processo nº 148127/2025-8) pela bolsa de iniciação científica e ao ITA pelo suporte institucional.

REFERÊNCIAS BIBLIOGRÁFICAS

- BOURNE, R. R. A. et al. Lancet Global Health, 2021.
- RIBEIRO, G. B. et al. ABO Online, 2021.
- WHO. World report on vision. 2019.
- SILVA, N. A. de O. et al. arXiv, 2024.