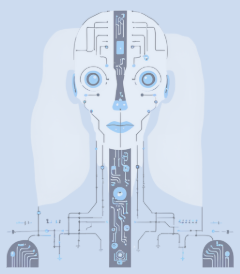
****TJCE****

***Coordenadoria de Pesquisa e Desenvolvimento de Soluções de IA***

****Relatório Técnico****

**Comparativo de Soluções de Transcrição de Áudio**

**(Azure Speech vs Whisper local)**



**Preparado por:**

**Rodrigo Garcia**

Sumário

[1. Objetivo 3](#__RefHeading___Toc1373_654969961)

[2. Metodologia dos Testes 3](#__RefHeading___Toc1375_654969961)

[2.1 Fontes de Áudio Utilizadas 3](#__RefHeading___Toc1377_654969961)

[2.2 Critérios de Avaliação 3](#__RefHeading___Toc1379_654969961)

[3. Implementação da POC de testes 4](#__RefHeading___Toc1381_654969961)

[4. Resultados dos Testes 5](#__RefHeading___Toc1383_654969961)

[4.1 Azure AI Speech – Transcrição em Lote 5](#__RefHeading___Toc1385_654969961)

[4.2 Whisper (rodando em Cluster com GPU) 5](#__RefHeading___Toc1387_654969961)

[4.3 Comparativo de desempenho 6](#__RefHeading___Toc1389_654969961)

[5. Comparativo Técnico 7](#__RefHeading___Toc1391_654969961)

[6. Estimativa de Custos 7](#__RefHeading___Toc1393_654969961)

[6.1 Azure AI Speech 7](#__RefHeading___Toc1395_654969961)

[6.2 Whisper Local 7](#__RefHeading___Toc1397_654969961)

[6.3 Estimativa de Custos de Armazenamento na Nuvem da Azure 8](#__RefHeading___Toc1399_654969961)

[7. Recomendação Final 8](#__RefHeading___Toc1401_654969961)

## ****1. Objetivo****

Este relatório visa apresentar os resultados dos testes comparativos entre as soluções de transcrição:

* **Azure AI Speech (Transcrição em Lote)**
* **Whisper (modelo open-source rodando em cluster com GPU)**

O objetivo é avaliar a qualidade, desempenho, limitações técnicas e custos envolvidos, auxiliando na escolha da melhor arquitetura para transcrição e diarização de áudios de reuniões e eventos corporativos.

## ****2. Metodologia dos Testes****

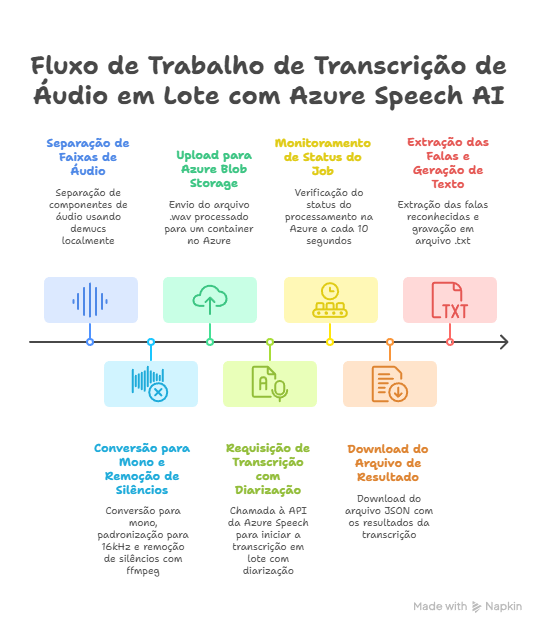
### 2.1 Fontes de Áudio Utilizadas

* Áudios reais de reuniões com múltiplos participantes
* Diversos formatos (.mp4, .asf), durações (curtos e longos) e qualidades (compressão, ruídos)

### 2.2 Critérios de Avaliação

* **Precisão da transcrição** (palavras corretas, fluência)
* **Diarização** (atribuição correta por locutor)
* **Tempo de processamento**
* **Facilidade de integração**
* **Custo operacional estimado**

## ****3. Implementação da POC de testes****



## ****4. Resultados dos Testes****

### 4.1 Azure AI Speech – Transcrição em Lote

| Critério | Observações |
| --- | --- |
| Qualidade da transcrição | Boa em ambientes claros; dificuldades com sobreposição e sotaques regionais. |
| Diarização | Funcional, mas falha em separações próximas ou com mais de 3 interlocutores. |
| Tempo de resposta | ~1/3 da duração do áudio (ótimo para lotes assíncronos). |
| Integração via API | Bem documentada; ausência de callbackUrl, exige implementação de webhook para notificação de conclusão. |
| Custo estimado | R$ 1,277 por hora de áudio (aproximadamente), com variações por região e uso de modelos customizados. |
| Limitações | Não possui controle refinado de segmentação nem acesso fácil aos logits. |

### 4.2 Whisper (rodando em Cluster com GPU)

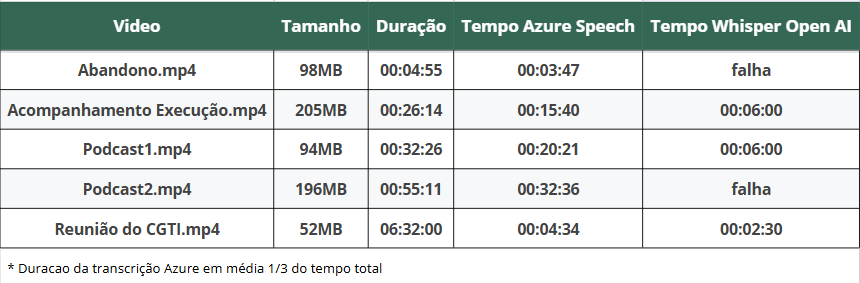
| Critério | Observações |
| --- | --- |
| Qualidade da transcrição | Muito boa, especialmente com large-v3; superior em sotaques e sobreposições. |
| Diarização | Necessita integração com modelos externos (ex: pyannote-audio). |
| Tempo de resposta | Depende do hardware; |
| Integração via API | Flexível; requer estruturação personalizada para escalabilidade. |
| Custo estimado | (infraestrutura) ?? |
| Limitações | Maior complexidade de operação e manutenção técnica, devido aos vários modelos e bibliotecas utilizados no processo. |

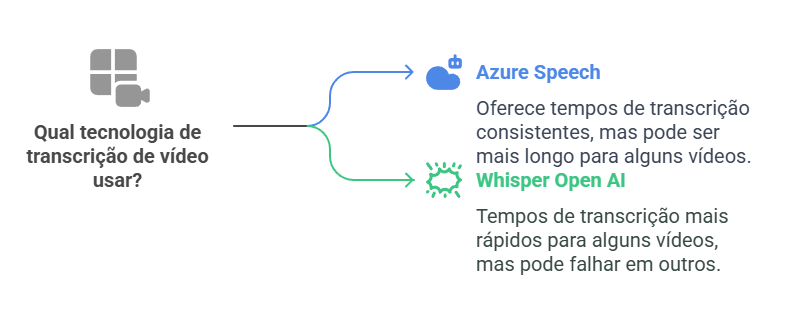
### 4.3 Comparativo de desempenho

**Testes efetuados em vários arquivos nas duas soluções, porém com condições não equivalentes.**

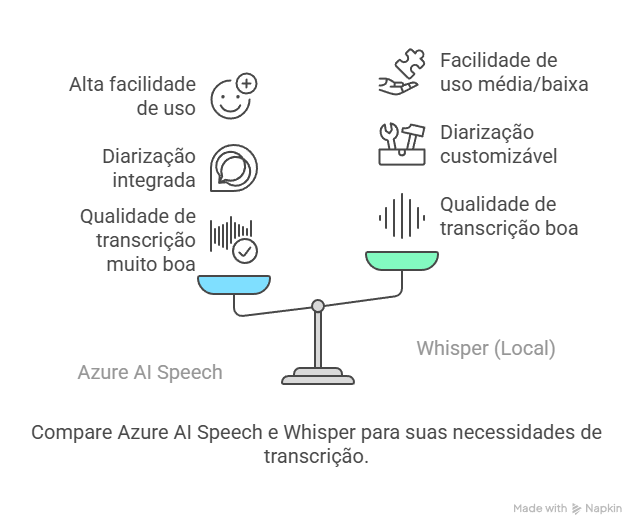
**a) Azure Speech : melhorias e conversões de aúdio efetuadas em máquina local, com limitações de hardware, e após tratamento processamento realizada na nuvem da Azure. O tempo registrado na tabela a seguir é composto pelos processamentos da máquina local + transcrição na Azure.**

**2) Whisper : utilizado o ambiente de teste da talia para transcrição dos arquivos. Em dois deles não foi possível colher o resultado pois foi apresentada falha no processo.**





## ****5. Comparativo Técnico****



## ****6. Estimativa de Custos****

### 6.1 Azure AI Speech

**Custos estimados por hora (transcrição + diarização)**

* Transcrição em lote (com diarização): R$1,325 (transcrição) + R$ 1,729 (diarização) por hora de áudio. Total = R$ 3,05
* Custo de armazenamento (BLOB Storage) (seção 5.3)

### 6.2 Whisper Local

* Infraestrutura: cluster com GPUs (ex.: 4x A10 = R$ X mil/mês)
* Licença: Open-source (sem custo)

### 6.3 Estimativa de Custos de Armazenamento na Nuvem da Azure

Para armazenar os áudios transcritos e os resultados das transcrições, consideramos o uso do **Azure Blob Storage**. Os custos variam conforme o volume de dados, a frequência de acesso e o nível de redundância escolhido.

**Custos estimados por armazenamento em GB/mês (Pay-as-you-go / 1 USD = 5.674 BRL ):**

* **Hot Tier (acesso frequente):** R$ 0,1850 por GB
* **Cool Tier (acesso esporádico):** R$ 0,10043 por GB
* **Archive Tier (acesso raro):** R$ 0,011816 por GB

**Exemplo de custo mensal para armazenamento de 1 TB de dados:**

* **Hot Tier:** R$ 189,44,00
* **Cool Tier:** R$ 102,84
* **Archive Tier:** R$ 12,09

## ****7. Recomendação Final****

A escolha entre as soluções deve considerar:

* **Volume de dados processados mensalmente**
* **Nível de controle exigido sobre os dados e modelos**
* **Equipe técnica disponível para operação da solução própria**
* **Orçamento disponível para investimento inicial e manutenção**