

EXPLORAÇÃO DA VOZ COMO MECÂNICA INTERATIVA EM JOGOS DIGITAIS: IMPACTOS NA EXPERIÊNCIA DO USUÁRIO E JOGABILIDADE

Elivan Ricardo Spiecker¹, Nicolás Sartori Emer², Edson Moacir Ahlert³

Resumo: Este trabalho investiga a interação por voz como uma mecânica inovadora em jogos digitais, com foco em RPGs 2D. A proposta utiliza o OpenAI Whisper para transformar comandos de voz em ações no jogo, criando uma experiência mais imersiva e natural para os jogadores. Entre os desafios técnicos abordados estão a adaptação a diferentes sotaques e a redução de ruídos, resolvidos com técnicas de pré-processamento acústico e treinamento contínuo do sistema. Além disso, o protótipo foi implementado na plataforma Unity, escolhida por sua flexibilidade e integração com bibliotecas externas. Os resultados iniciais indicam o potencial dessa abordagem para transformar a jogabilidade, ampliando a acessibilidade e promovendo maior imersão. Ao optar por uma solução de código aberto, o estudo demonstra que é possível alcançar resultados significativos mesmo em contextos com restrições orçamentárias, contribuindo para a democratização do uso de interfaces interativas baseadas em voz.

Palavras-chave: interação por voz; jogos digitais; reconhecimento de fala; inteligência artificial.

1 INTRODUÇÃO

A interação por voz em jogos eletrônicos tem se destacado como uma ferramenta promissora para aumentar a imersão e a interatividade, especialmente em gêneros que dependem de profundidade narrativa, como os Role-Playing Games (RPGs). Esses jogos são conhecidos por proporcionar uma experiência envolvente e dinâmica, na qual o jogador assume o papel de um

1 Estudante de Engenharia de Software - Universidade do Vale do Taquari - Univates. E-mail: elivan.spiecker@universo.univates.br

2 Estudante de Engenharia de Software - Universidade do Vale do Taquari - Univates. E-mail: nicolas.emer@universo.univates.br

3 Professor da Universidade do Vale do Taquari - Univates. E-mail: edsonahlert@univates.br

personagem dentro de um universo fictício e interage com outros personagens e elementos desse ambiente.

A interação ocorre por meio de comandos baseados em menus ou controle manual, mas, com os avanços em reconhecimento de fala e processamento de linguagem natural, a possibilidade de utilizar comandos de voz surge como uma forma inovadora de melhorar a experiência do usuário.

A pesquisa de DigitalizaPro (2024) aponta que, apesar dos avanços no uso de inteligência artificial e aprendizado de máquina, o reconhecimento de voz ainda enfrenta dificuldades significativas, como variações de sotaques e ruídos de fundo, que impactam diretamente na precisão e no desempenho.

A tecnologia de reconhecimento de fala, essencial para a proposta deste trabalho, enfrenta diversos desafios, especialmente em ambientes interativos, como jogos. De acordo com a MozBazar (2024), superar esses desafios é essencial para garantir que a interação entre humanos e sistemas de reconhecimento de fala seja natural e eficiente, contribuindo para ampliar o alcance e a usabilidade dessa tecnologia em diferentes contextos.

A pesquisa de Carter *et al.* (2015) destaca que, apesar do potencial para aumentar a imersão, a interação por voz em jogos ainda enfrenta obstáculos técnicos que comprometem a naturalidade das interações. Tonini (2023) também aponta que, embora a tecnologia de voz seja promissora, ela exige melhorias significativas para garantir uma resposta natural e fluida, principalmente quando aplicada em jogos com elementos de realidade virtual, onde a expectativa de imersão é ainda maior.

Além dos desafios técnicos, questões relacionadas à privacidade e ao uso ético de dados de voz também são levantadas por estudiosos como Zargham *et al.* (2024), que apontam a necessidade de regulamentações que protejam os dados dos usuários. A coleta e o uso de dados de voz em jogos demandam políticas claras de privacidade, especialmente em um contexto em que essas interações podem ser gravadas e processadas por sistemas de inteligência artificial. O uso ético dessas informações é um fator determinante para a aceitação dessas tecnologias pelos jogadores.

Este trabalho tem como objetivo desenvolver um protótipo de jogo RPG 2D que utiliza comandos de voz para interagir com NPCs e realizar ações dentro do ambiente de jogo. O protótipo será desenvolvido na plataforma Unity, com a integração do sistema OpenAI Whisper para conversão de fala em texto. A hipótese central é que a interação por voz pode aumentar significativamente a imersão do jogador, oferecendo uma experiência mais dinâmica e natural.

Ao longo do desenvolvimento, será analisada a eficiência do sistema de reconhecimento de fala em termos de precisão, adaptação a diferentes sotaques e desempenho em ambientes com ruído. Além disso, será realizada uma comparação entre o OpenAI Whisper e outras soluções de reconhecimento de

voz, como o Google Speech-to-Text e a Amazon Polly, considerando aspectos como custo, precisão e facilidade de integração.

2 REFERENCIAL TEÓRICO

O desenvolvimento de soluções tecnológicas que envolvem a interação por voz em jogos digitais está alicerçado em uma série de avanços nas áreas de reconhecimento de fala, inteligência artificial e desenvolvimento de jogos. Para criar um jogo RPG com comandos de voz, é fundamental entender as bases teóricas e tecnológicas que possibilitam essa integração.

Neste capítulo, serão discutidos os trabalhos anteriores que investigam a interação por voz em jogos, as tecnologias e frameworks que permitem o reconhecimento de fala, além das metodologias de desenvolvimento de software que são aplicadas na criação dessas soluções.

2.1 Reconhecimento de Fala e Interação em Jogos

O reconhecimento de fala é a tecnologia que possibilita que máquinas interpretem comandos verbais emitidos por humanos. Ao longo dos últimos anos, essa tecnologia avançou consideravelmente devido ao desenvolvimento de algoritmos baseados em redes neurais profundas e técnicas de processamento de linguagem natural (NLP). No contexto dos jogos eletrônicos, o reconhecimento de fala tem sido explorado como uma forma de aumentar a imersão e facilitar a interação entre jogadores e o ambiente do jogo.

De acordo com Carter *et al.* (2015), a interação por voz pode alinhar a identidade do jogador com o personagem do jogo, oferecendo uma experiência mais natural e fluida. No entanto, os autores também apontam para a existência de desafios, como a precisão da tecnologia e o desconforto social que pode surgir ao utilizar comandos de voz em jogos.

Pesquisas anteriores, como a de Allison, Gibbs e Carter (2016), traçaram um histórico da utilização de comandos de voz em jogos, desde seus primeiros experimentos na década de 1960 até a popularização mais recente com o uso de dispositivos como o Microsoft Kinect e o Amazon Echo. Esses dispositivos abriram caminho para uma nova geração de jogos que exploram a fala como uma forma de controle, permitindo que os jogadores realizem ações por meio de comandos verbais. No entanto, apesar dos avanços tecnológicos, os jogos baseados em voz ainda enfrentam limitações significativas, especialmente no que diz respeito à naturalidade das respostas dos NPCs e à adaptação do sistema de reconhecimento de fala a diferentes contextos linguísticos e culturais.

No âmbito dos RPGs, a imersão é um dos fatores mais importantes. O estudo de Tonini (2023) explora como a interação por voz pode melhorar a experiência de imersão em jogos de realidade virtual, enfatizando a importância de uma resposta natural dos NPCs. Segundo os autores, uma das principais dificuldades é garantir que o reconhecimento de fala funcione de forma

precisa em tempo real, sem comprometer a fluidez do jogo. Neste contexto, tecnologias como o OpenAI Whisper, utilizadas no presente trabalho, surgem como alternativas promissoras por sua alta taxa de precisão e capacidade de adaptação em diferentes cenários.

2.2 Tecnologias de Reconhecimento de Fala

As soluções de reconhecimento de fala baseadas em redes neurais profundas, como o OpenAI Whisper, utilizam grandes quantidades de dados de voz para treinar seus modelos, permitindo que eles reconheçam com precisão uma ampla gama de sotaques e variações linguísticas.

O Whisper, tecnologia de código aberto desenvolvida pela OpenAI, oferece uma solução eficiente para a transcrição de fala em texto, evitando os custos associados a APIs comerciais, como o Google Speech-to-Text e a Amazon Polly. A escolha do Whisper para o desenvolvimento do jogo proposto foi motivada não apenas pela sua precisão, mas também pela flexibilidade que a ferramenta oferece em termos de adaptação e integração com diferentes plataformas, como a Unity.

O Google Speech-to-Text e a Amazon Polly também são amplamente utilizados em soluções de reconhecimento de fala, mas apresentam algumas limitações em relação ao custo e à acessibilidade. Essas tecnologias, embora extremamente precisas, cobram por requisição de uso, o que pode tornar a implementação de jogos com comandos de voz financeiramente inviável, especialmente em projetos independentes ou com orçamento limitado.

Além disso, a dependência de servidores externos para processar comandos de voz pode introduzir latências indesejadas, comprometendo a experiência do jogador em jogos que demandam resposta imediata.

Pesquisas como as de Filimon, Iftene e Trandabăț (2019) também destacam a importância de adaptar a tecnologia de reconhecimento de fala ao contexto do jogo. Em seu estudo sobre jogos educacionais com interação por voz, os autores sugerem que o uso de assistentes de voz, como a Alexa da Amazon, pode facilitar o aprendizado em ambientes interativos. A aplicação dessa tecnologia em jogos educacionais e de entretenimento demonstra a versatilidade e o potencial do reconhecimento de fala para transformar a forma como os usuários interagem com os sistemas computacionais.

2.3 Unity como Plataforma de Desenvolvimento

O desenvolvimento de jogos com comandos de voz requer uma engine flexível e poderosa o suficiente para lidar com a integração de diferentes tecnologias. A Unity foi escolhida como a plataforma principal para o desenvolvimento deste projeto devido à sua popularidade na indústria de jogos e à sua capacidade de integração com APIs externas, como o OpenAI Whisper. A Unity é uma ferramenta amplamente utilizada por desenvolvedores de jogos

de diferentes níveis de experiência, desde iniciantes até profissionais, e oferece suporte robusto para jogos 2D e 3D, além de uma extensa documentação e uma comunidade ativa de desenvolvedores.

De acordo com Córdula (2022), a Unity se destaca pela facilidade com que permite a criação de protótipos rápidos, o que é essencial em projetos que envolvem a experimentação com novas tecnologias, como o reconhecimento de fala. A capacidade da Unity de integrar scripts em C# e utilizar bibliotecas externas para expandir suas funcionalidades é um de seus maiores pontos fortes. Isso permite que os desenvolvedores adaptem a engine às necessidades específicas de cada projeto, garantindo a flexibilidade necessária para implementar um sistema de comandos de voz eficiente.

Além disso, a Unity possui uma vasta gama de recursos pré-configurados, como sistemas de animação, física e gerenciamento de interface de usuário, o que facilita a criação de jogos que ofereçam uma experiência visualmente rica e interativa. No contexto do desenvolvimento de um jogo RPG com comandos de voz, a Unity permite a implementação de diálogos dinâmicos com NPCs, onde os jogadores podem interagir verbalmente, aumentando a imersão sem sacrificar a jogabilidade.

2.4 Metodologias de Desenvolvimento de Software para Jogos

A criação de um jogo interativo, especialmente um que utilize tecnologias de ponta como o reconhecimento de fala, exige o uso de metodologias de desenvolvimento de software que garantam a organização do projeto e a entrega de um produto final de alta qualidade. No desenvolvimento de jogos, metodologias ágeis, como o Scrum e o Kanban, são frequentemente empregadas para organizar as fases do projeto, do planejamento e prototipagem à implementação e testes. Essas metodologias permitem uma maior flexibilidade, o que é fundamental quando se trabalha com tecnologias emergentes que podem exigir ajustes frequentes durante o desenvolvimento.

De acordo com Alves (2019), o desenvolvimento ágil é especialmente útil em projetos de jogos que envolvem prototipagem rápida e iteração contínua, como é o caso deste projeto, onde o sistema de reconhecimento de fala precisa ser testado e ajustado conforme os feedbacks dos usuários. A utilização de sprints curtos e a constante revisão das funcionalidades implementadas permite que a equipe de desenvolvimento identifique problemas e implemente soluções de forma ágil, garantindo que o jogo final atenda às expectativas em termos de usabilidade e desempenho.

A abordagem baseada em User-Centered Design (UCD) também é relevante neste contexto, já que o foco principal do jogo é proporcionar uma experiência imersiva e intuitiva para os jogadores. O UCD coloca o jogador no centro do processo de desenvolvimento, garantindo que todas as decisões de

design e implementação sejam feitas com base nas necessidades e preferências do usuário final.

3 PROCEDIMENTOS METODOLÓGICOS

Para a criação de um produto de software com enfoque em comandos de voz para interação em jogos RPG 2D, foram estabelecidos procedimentos metodológicos específicos, que abarcam desde a descrição e caracterização da pesquisa até a análise dos resultados obtidos com base no desempenho e usabilidade do sistema. Este processo metodológico foi elaborado com o objetivo de garantir que o desenvolvimento, implementação e validação do software fossem conduzidos de forma a atender às expectativas de imersão e interatividade definidas no escopo inicial.

3.1 Descrição e Caracterização da Pesquisa

A pesquisa aqui descrita foi estruturada para investigar e desenvolver uma interface interativa por comandos de voz em jogos digitais, mais especificamente no gênero RPG 2D. Diante do crescente interesse na utilização de voz como interface em jogos, especialmente RPGs, a pesquisa objetivou aprimorar a imersão dos jogadores ao possibilitar uma comunicação direta com NPCs (personagens não jogáveis) através de comandos verbais. Estudos anteriores, como os de Carter *et al.* (2015), ressaltam que a interação por voz não apenas facilita a navegação, mas também transforma a experiência de jogo, criando um ambiente mais responsivo e imersivo.

Dessa forma, a pesquisa caracteriza-se como aplicada e experimental, pois visa a aplicação prática de tecnologias de reconhecimento de fala no desenvolvimento de um protótipo de jogo. A abordagem experimental foi essencial para permitir que os testes e ajustes fossem realizados com flexibilidade, de modo a avaliar a viabilidade técnica e as possibilidades de aprimoramento dos comandos de voz em um ambiente de jogo dinâmico.

3.2 Concepção e Planejamento da Solução

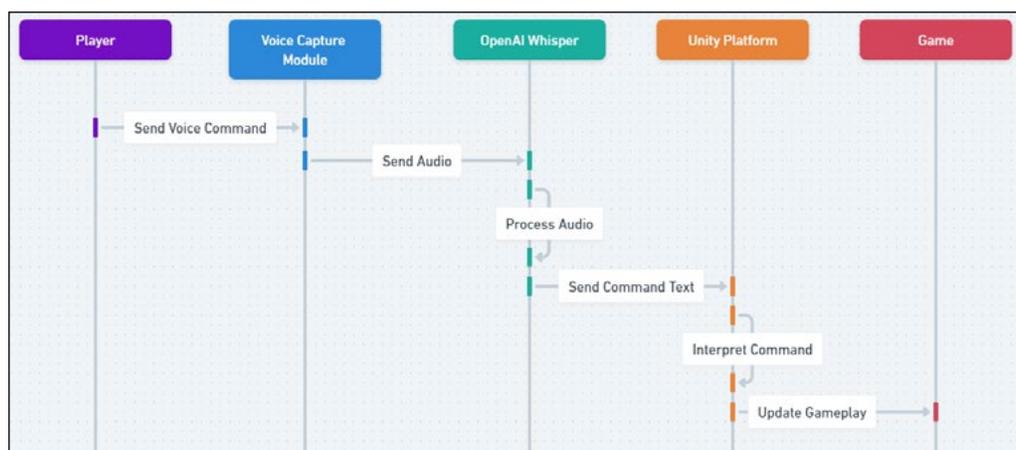
Para a concepção e planejamento do software, foi definido um conjunto de ferramentas e frameworks que atendessem tanto aos requisitos de funcionalidade quanto aos de viabilidade técnica e econômica. A escolha das ferramentas foi orientada pela necessidade de alta precisão no reconhecimento de voz, baixa latência e flexibilidade para integração em um ambiente de desenvolvimento de jogos.

A Unity foi selecionada como plataforma principal de desenvolvimento devido à sua popularidade e à vasta gama de recursos disponíveis, que permitem a criação de protótipos rápidos e de alta qualidade. Além disso, a Unity oferece uma integração facilitada com bibliotecas externas, como

o OpenAI Whisper, possibilitando a implementação de funcionalidades avançadas de reconhecimento de fala. Conforme Córdula (2022), a Unity é amplamente utilizada para projetos que requerem integração de múltiplas tecnologias e experimentação contínua, características centrais para o sucesso desta pesquisa.

O OpenAI Whisper foi escolhido como sistema de reconhecimento de fala, principalmente pela sua adaptabilidade a diferentes sotaques e sua eficiência em ambientes com ruído moderado. Essa tecnologia é de código aberto, o que a torna uma alternativa acessível a sistemas comerciais que requerem pagamento por requisição de uso, como o Google Speech-to-Text e a Amazon Polly. O OpenAI Whisper foi configurado para converter a fala dos jogadores em texto, permitindo uma comunicação natural e fluida com os NPCs, elevando a experiência de imersão proposta.

Figura 1 - Diagrama de sequência dos componentes



Fonte: Dos autores (2024).

O desenvolvimento do sistema foi planejado em etapas, com foco na integração eficiente do OpenAI Whisper à plataforma Unity. Um dos elementos centrais dessa fase foi o uso de um diagrama de sequência (Figura 1), que ilustra a interação entre os componentes principais do sistema, desde a captura de comandos de voz pelo jogador até a execução das ações no ambiente do jogo.

O diagrama apresenta, de forma estruturada, o fluxo de dados entre o jogador, o sistema de reconhecimento de fala e os componentes do jogo. Ele evidencia como os comandos de voz são processados em tempo real, passando por etapas de conversão de fala para texto, interpretação do comando e execução da ação correspondente.

Essa visualização foi essencial para identificar pontos críticos no fluxo de processamento, permitindo ajustes na arquitetura do sistema para minimizar

latências e garantir uma resposta ágil e precisa. A utilização do diagrama também contribuiu para alinhar os esforços de design e desenvolvimento, assegurando que todos os aspectos do sistema fossem integrados de maneira coesa e funcional.

3.3 Desenvolvimento e Implementação da Solução

O desenvolvimento da solução foi dividido em três fases principais: análise de requisitos, design de sistema e desenvolvimento do código. Essa divisão possibilitou uma abordagem mais detalhada e focada em cada etapa do processo, garantindo que cada aspecto do software fosse implementado conforme os objetivos iniciais do projeto.

1. **Análise de Requisitos:** A fase de análise de requisitos envolverá a definição das funcionalidades essenciais do sistema de comandos de voz, incluindo precisão, baixa latência e responsividade. Serão realizados levantamentos junto a potenciais usuários do jogo para identificar as principais necessidades e expectativas relacionadas à interação por voz. Com isso, será possível delinear os requisitos funcionais (como comandos específicos de voz) e os requisitos não funcionais (como o desempenho em ambientes com ruído e a adaptabilidade a diferentes sotaques).
2. **Design do Sistema:** Com os requisitos definidos, será desenvolvida a arquitetura do sistema, que integrará o Whisper à plataforma Unity. O design do sistema focará na construção de uma interface que possibilite a tradução direta de comandos verbais em ações dos personagens, promovendo uma interação intuitiva e imersiva. Durante essa fase, o sistema será estruturado para reconhecer comandos verbais específicos que desencadeiem ações dos NPCs e para lidar com variações na pronúncia e entonação dos jogadores, aspectos apontados como críticos por Allison, Carter e Gibbs (2016) em seus estudos sobre interações em jogos digitais.

Figura 2 - Tela de Menu Inicial



Fonte: Dos autores (2024).

3. Desenvolvimento do Código: A fase de desenvolvimento do código envolverá a implementação do protótipo utilizando scripts em C#, uma linguagem amplamente suportada pela Unity e adequada para manipulação de dados de voz. Nessa etapa, o Whisper será configurado para capturar comandos de voz em tempo real, convertendo-os em texto e integrando-os diretamente ao ambiente do jogo. Serão implementadas rotinas para otimizar o tempo de resposta do sistema e garantir que o comando verbal seja executado sem atrasos perceptíveis, proporcionando uma experiência fluida para o jogador.

3.4 Validação do Software

A validação do software será realizada em três frentes principais: a criação de protótipos e a condução de testes com usuários em ambientes

simulados. Essas etapas foram cruciais para verificar a eficiência do sistema de reconhecimento de fala e sua aplicabilidade em situações reais de uso.

1. Prototipagem: O protótipo do jogo será desenvolvido em ciclos iterativos, com o objetivo de testar e refinar a implementação dos comandos de voz. Cada nova versão do protótipo passará por uma análise rigorosa de desempenho e precisão do reconhecimento de fala, o que permitirá identificar e corrigir eventuais problemas de interpretação ou latência no processamento dos comandos.

Figura 3 - Protótipo de ambientação, para testes de voz



Fonte: Dos autores (2024).

2. Testes com Usuários: Serão conduzidos testes com um grupo de jogadores com diferentes características vocais, incluindo variações de sotaques e timbres. Os testes serão realizados em ambientes com diferentes níveis de ruído, simulando condições reais de uso. A análise do feedback dos usuários, combinada aos dados de precisão do reconhecimento de fala, indicará se o Whisper será eficaz para comandos em ambientes com ruído moderado, além de apontar possíveis limitações em ruídos intensos. Esses testes confirmarão a viabilidade do sistema e orientarão possíveis melhorias a serem implementadas em versões futuras.
3. Ambientes Simulados: Para validar a responsividade e usabilidade do software em diferentes contextos, serão criados ambientes simulados que replicarão cenários reais de jogo, como salas de conferência e

áreas públicas com níveis variados de interferência sonora. Essa abordagem permitirá avaliar a adaptabilidade do Whisper a ruídos de fundo e condições acústicas desafiadoras, o que será fundamental para entender o potencial de uso do sistema em diferentes cenários.

4 RESULTADOS ESPERADOS E DISCUSSÃO

A interação por voz em jogos digitais apresenta grande potencial para transformar a experiência do usuário, oferecendo uma forma mais intuitiva e natural de comunicação com os elementos do jogo. Este projeto visa explorar essa possibilidade em um RPG 2D, utilizando o OpenAI Whisper para integrar comandos verbais às mecânicas do jogo. Essa abordagem não só busca aumentar a imersão do jogador, mas também simplificar interações complexas com NPCs e ambientes virtuais.

Embora promissora, a implementação de comandos de voz enfrenta desafios significativos, como a adaptação a diferentes sotaques e a redução de ruídos. Como destacado por Tonini (2023), a imersão em jogos depende diretamente da capacidade do sistema de responder de forma natural e fluida, o que reforça a importância de tecnologias como o Whisper para alcançar essa experiência. A adoção de técnicas de pré-processamento acústico e treinamento contínuo são estratégias fundamentais para superar essas barreiras e garantir uma interação eficiente.

Os resultados esperados incluem avanços tanto no campo técnico quanto no design centrado no usuário, com foco na acessibilidade e na usabilidade. De acordo com Carter *et al.* (2015), a identidade do jogador é fortalecida quando as interações com o jogo são alinhadas às suas expectativas, evidenciando que interfaces de voz podem substituir métodos tradicionais de controle por abordagens mais orgânicas e envolventes. Assim, o projeto busca demonstrar como a interação verbal com NPCs pode proporcionar uma experiência de jogo mais fluida, especialmente em um cenário onde ferramentas acessíveis, como o OpenAI Whisper, se mostram eficientes e econômicas.

4.1 Resultados obtidos até o momento

Até o momento, foram desenvolvidos conceitos visuais essenciais, como o design das telas do menu principal, sistema de inventário e ambientação. Esses materiais iniciais ajudam a alinhar os elementos do jogo aos objetivos de interação planejados. Conforme apontado por Zargham *et al.* (2024), é crucial que tecnologias de interação por voz sejam implementadas de forma ética, protegendo os dados dos usuários e garantindo um design centrado no jogador. Essa perspectiva é incorporada no projeto, que valoriza a privacidade e busca criar uma experiência imersiva e segura.

Figura 4 - Protótipo de inventário



Fonte: Dos autores (2024).

Embora o progresso inicial tenha se concentrado em elementos conceituais, esses resultados fornecem uma base sólida para as próximas etapas, que incluirão testes de jogabilidade e integração dos comandos de voz. Dessa forma, o projeto avança em alinhamento com seus objetivos principais, enquanto contribui para o desenvolvimento de interfaces verbais replicáveis e acessíveis no campo dos jogos digitais.

4.2 Desafios e Soluções Propostas

A adaptação linguística e o desempenho do sistema em ambientes variados são aspectos cruciais para o sucesso do projeto de interação por voz em RPGs. Entre os principais desafios enfrentados, destacam-se as diferenças de sotaques regionais e pronúncias, que podem impactar a eficácia da detecção de fala. Para mitigar esses problemas, propõe-se o treinamento contínuo do sistema utilizando um conjunto diversificado de dados de voz. Além disso, a

implementação de filtros especializados ajudará a refinar a interpretação de comandos, melhorando a precisão mesmo diante de variações linguísticas.

Outro desafio identificado está relacionado à presença de ruídos ambientais intensos, que podem comprometer a precisão do sistema, e à latência nas respostas, que prejudica a experiência do usuário. Para superar essas dificuldades, sugere-se a aplicação de técnicas avançadas de pré-processamento acústico, como filtros de ruído e normalização de áudio, além da otimização dos scripts de processamento local, buscando reduzir atrasos e garantir um funcionamento mais ágil e responsivo.

4.3 Próximos Passos

O projeto avança para uma fase de aprimoramento do protótipo, que incluirá melhorias baseadas nos feedbacks coletados durante os testes iniciais. Essa etapa envolverá também a ampliação dos testes em cenários acústicos variados, visando avaliar o desempenho do sistema em condições reais de uso.

Além disso, planeja-se realizar testes ampliados com grupos diversificados de usuários, considerando perfis linguísticos distintos e diferentes níveis de experiência com tecnologias de interação por voz. Essa abordagem permitirá uma análise mais abrangente da usabilidade e eficiência do sistema.

Por fim, será iniciada a elaboração de publicações científicas detalhando os avanços, desafios e soluções encontradas ao longo do desenvolvimento. Esses artigos têm como objetivo compartilhar os resultados com a comunidade acadêmica e a indústria, contribuindo para o avanço das pesquisas e aplicações em interação por voz.

5 CONSIDERAÇÕES FINAIS

O avanço na integração de comandos de voz em jogos digitais, como explorado neste projeto, representa uma inovação significativa na forma como jogadores interagem com ambientes virtuais. Este trabalho focou na implementação de um sistema funcional de reconhecimento de fala utilizando o OpenAI Whisper em um RPG 2D, com o objetivo de criar uma experiência de jogo mais intuitiva, natural e imersiva. A pesquisa revelou que, embora os desafios técnicos e de design sejam substanciais, as soluções propostas têm o potencial de transformar a interação entre jogador e jogo, promovendo um novo padrão de jogabilidade em um campo altamente dinâmico.

Os principais desafios enfrentados serão relacionados à adaptação linguística e à precisão em ambientes com ruído, como apontado por Carter *et al.* (2015), que destacam as limitações técnicas que podem comprometer a fluidez das interações em jogos digitais. Para mitigar esses problemas, serão propostas soluções como o treinamento contínuo com dados diversificados e a aplicação de técnicas de pré-processamento acústico, que permitem refinar

a interpretação de comandos verbais. Essas estratégias não apenas aumentam a precisão do sistema, mas também reforçam a capacidade de adaptação a diferentes contextos de uso, alinhando-se às perspectivas discutidas por Tonini (2023) sobre a importância de respostas naturais para garantir imersão em jogos.

Outro aspecto relevante abordado foi a ética e a privacidade no uso de dados de voz, um tema amplamente discutido por Zargham *et al.* (2024). O projeto irá incorporar diretrizes claras para assegurar que os dados coletados sejam tratados de forma segura, respeitando a privacidade dos usuários. Essa abordagem não só aumenta a confiança dos jogadores, mas também estabelece uma base sólida para o uso responsável de tecnologias de reconhecimento de fala em jogos e outras aplicações interativas.

Os resultados obtidos até o momento, como a concepção de elementos visuais fundamentais para o jogo, mostram que o desenvolvimento inicial está alinhado com os objetivos propostos. Embora funcionalidades avançadas ainda não tenham sido implementadas, os conceitos desenvolvidos fornecem uma base sólida para as próximas etapas. A criação de interfaces intuitivas e alinhadas com a integração verbal reforça a importância de um design centrado no usuário, conforme destacado por estudos prévios no campo.

Este trabalho também destaca o potencial do OpenAI Whisper como uma ferramenta de reconhecimento de fala acessível e eficiente. Ao optar por uma tecnologia de código aberto, o projeto demonstra que é possível alcançar resultados significativos mesmo em contextos com restrições de orçamento. Essa escolha não só valida a viabilidade técnica do sistema, mas também contribui para a democratização do uso de interfaces interativas baseadas em voz, promovendo maior acessibilidade tecnológica.

Por fim, o impacto esperado do projeto transcende os limites dos jogos digitais. As aplicações potenciais de comandos de voz em outros campos, como educação, saúde e atendimento ao cliente, abrem novas perspectivas para a interação humano-computador. A continuidade do trabalho incluirá a ampliação de testes com grupos diversificados de usuários, o refinamento do protótipo e a disseminação dos resultados na forma de publicações científicas. Essas etapas não apenas consolidarão os avanços obtidos, mas também promoverão um diálogo mais amplo sobre as possibilidades e os desafios das interfaces verbais.

O projeto, portanto, representa um marco inicial na exploração de uma tecnologia que tem o poder de redefinir a jogabilidade e a interação em ambientes digitais. Ao abordar de forma inovadora e ética os desafios técnicos e sociais envolvidos, ele contribui para o avanço do design centrado no usuário e para a expansão do uso de tecnologias de voz em um mundo cada vez mais conectado.

REFERÊNCIAS

ALVES, William Pereira. **Unity design e desenvolvimento de jogos**. Rio de Janeiro: Alta Books, 2019.

ALLISON, Fraser; CARTER, Marcus; GIBBS, Martin. **A history of voice interaction in games**. In: DIGRA/FDG '16 – Proceedings of the First International Joint Conference of DiGRA and FDG, 2016. Disponível em: <https://dl.digra.org/index.php/dl/article/download/887/887>. Acesso em: 30 set. 2024.

CARTER, Marcus; ALLISON, Fraser; DOWNS, John; GIBBS, Martin. **Player Identity Dissonance and Voice Interaction in Games**. In: CHI PLAY 2015. Proceedings... London: ACM, 2015. p. 1-9. Disponível em: <https://rest.neptune-prod.its.unimelb.edu.au/server/api/core/bitstreams/86082fba-7d4e-5445-b4e2-e4fbdd8efba6/content>. Acesso em: 30 set. 2024.

CÓRDULA, Rodrigo Beltrão de Lucena. **Criando Games Com Unity - Direto ao Ponto**. Ciência Moderna, 2022.

DIGITALIZAPRO. **Reconhecimento de voz: tecnologia e aplicações**. DigitalizaPro, 2024. Disponível em: <https://digitalizapro.com.br/glossario/reconhecimento-de-voz-tecnologia-e-aplicacoes>. Acesso em: 19 nov. 2024.

FILIMON, Marta; IFTENE, Adrian; TRANDABĂȚ, Diana. **Bob - A General Culture Game with Voice Interaction**. Procedia Computer Science, [S.l.], v. 159, p. 323–332, 2019. Disponível em: <https://doi.org/10.1016/j.procs.2019.09.187>. Acesso em: 12 nov. 2024.

MOZBAZAR. **Quebrando o silêncio: desafios no reconhecimento de fala**. MozBazar, 2024. Disponível em: <https://mozbazar.com/reconhecimento-de-fala-101-quebrando-o-silencio>. Acesso em: 19 nov. 2024.

TONINI, Leandro. **Talk to me, Hal: A Study of Player Experience and Interaction in a Voice Interaction VR Game Featuring AI-driven Non-player Characters**. 2024. Dissertação (Mestrado) – University of Twente, 2024. Disponível em: <https://essay.utwente.nl/98788/>. Acesso em: 30 set. 2024.

ZARGHAM, Nima; FRIEHS, Maximilian A.; TONINI, Leandro; ALEXANDROVSKY, Dmitry; RUTHVEN, Emma Grace; NACKE, Lennart E.; MALAKA, Rainer. **Let's Talk Games: An Expert Exploration of Speech Interaction with NPCs**. *International Journal of Human-Computer Interaction*, [S.l.], 2024. Disponível em: <https://doi.org/10.1080/10447318.2024.2338666>. Acesso em: 12 nov. 2024.