

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
CURSO DE GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO

Guilherme Henrique Galelli Christmann

**UM JOGO DE INTERAÇÃO HUMANO-ROBÔ PARA O ENSINO DE
LÍNGUA INGLESA**

Santa Maria, RS
2019

Guilherme Henrique Galelli Christmann

UM JOGO DE INTERAÇÃO HUMANO-ROBÔ PARA O ENSINO DE LÍNGUA INGLESA

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Computação da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Bacharel em Engenharia de Computação**.

ORIENTADOR: Prof. Rodrigo da Silva Guerra

Santa Maria, RS
2019

©2019

Todos os direitos autorais reservados a Guilherme Henrique Galelli Christmann. A reprodução de partes ou do todo deste trabalho só poderá ser feita mediante a citação da fonte.

Endereço: Rua Frederico Varaschini, 470, Ap. 101

Fone (0xx) 55 98120 9489; End. Eletr.: guichristmann@gmail.com

Guilherme Henrique Galelli Christmann

UM JOGO DE INTERAÇÃO HUMANO-ROBÔ PARA O ENSINO DE LÍNGUA INGLESA

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Computação da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Bacharel em Engenharia de Computação**.

Aprovado em 28 de janeiro de 2019:

Rodrigo da Silva Guerra, Dr. (UFSM)
(Presidente/Orientador)

Susana Cristina dos Reis, Dr. (UFSM)

Roseclea Duarte Medina, Dr. (UFSM)

Santa Maria, RS
2019

RESUMO

UM JOGO DE INTERAÇÃO HUMANO-ROBÔ PARA O ENSINO DE LÍNGUA INGLESA

AUTOR: Guilherme Henrique Galelli Christmann

ORIENTADOR: Rodrigo da Silva Guerra

Na sociedade contemporânea, a proficiência em língua inglesa é cada vez mais relevante. Sua padronização como língua de comunicação internacional e compartilhamento de conhecimento gera um grande interesse em seu aprendizado. Considerando esses fatores, existe uma busca por métodos mais eficazes de ensino e, atualmente, robôs mostram-se como ferramentas de grande potencial. Este trabalho apresenta um jogo de interação humano-robô (IHR) utilizando o robô **Beo** para o ensino de língua inglesa, baseando-se em princípios e diretrizes de IHR. Como resultado, é apresentado o princípio de uma plataforma de ensino, um robô fisicamente presente integrado com sistemas de reconhecimento de fala e objeto. As capacidades de IHR são demonstradas através de um jogo com objetos físicos.

Palavras-chave: Interação Humano-Robô. Robótica Humanóide. Ensino de Segunda Língua. Beo.

ABSTRACT

A HUMAN-ROBOT INTERACTION GAME FOR TEACHING ENGLISH LANGUAGE

AUTHOR: Guilherme Henrique Galelli Christmann

ADVISOR: Rodrigo da Silva Guerra

In the modern society, proficiency in the English language is increasingly relevant. Its standardization as the default language for international communication and information exchange generates a significant interest in the learning of English as a second language. With these factors in mind, there is a continuous search for more efficient teaching tools and, nowadays, robots are being established as a tool with great potential. This work presents a human-robot interaction (IHR) game, using the robot **Beo** to teach English language, relying on principles and guidelines of IHR. As a result, the principle of a teaching platform is presented, which integrates a physically present robot with speech and object recognition systems.

Keywords: Human-Robot Interaction. Humanoid Robotics. Second Language Acquisition. Beo.

LISTA DE FIGURAS

Figura 2.1 – Interesse no termo <i>gamification</i> desde 2010.	13
Figura 2.2 – Robovie de Kanda et al. (2004) interagindo com uma criança.	16
Figura 2.3 – Robô NIMA , utilizado por Alemi, Meghdari e Ghazisaedy (2014) para ensinar crianças a língua inglesa através de estórias.	17
Figura 3.1 – Robô Beo	18
Figura 3.2 – Foto de Tinker Board (à esquerda) e esquemático de entradas e saídas (à direita).	20
Figura 3.3 – Fotografia dos objetos físicos espalhados pelo chão.	22
Figura 3.4 – Placa de Desenvolvimento <i>NVIDIA Jetson TX2</i>	23
Figura 3.5 – Fluxo do jogo de interação com o Beo.	24
Figura 3.6 – Disposição dos jogadores e objetos físicos para o início do jogo.	25
Figura 3.7 – Ordem dos processos do sistema reconhecimento de fala completo.	27
Figura 3.8 – Entrada e saída de uma imagem com o modelo Tiny Yolo V3 treinado. ..	29
Figura 4.2 – Detecções positivas para 50 verbalizações para cada modelo do sistema de reconhecimento de fala em ambiente silencioso e barulhento.	36
Figura 4.3 – Detecções falsas em uma hora de gravação para cada modelo do sistema de reconhecimento (Etapa 1) de fala em ambientes silencioso e barulhento.	37
Figura 4.4 – Detecções positivas para 15 verbalizações para cada palavra do sistema de reconhecimento de fala (Etapa 3) em ambientes silencioso e barulhento.	38
Figura 4.5 – Gráfico do erro ao longo do treinamento para o conjunto de dados de treino.	40
Figura 4.6 – <i>AP</i> (Precisão Média) para cada classe de objetos no conjunto de dados de validação.	41
Figura 4.7 – <i>mAP</i> (Média das Precisões Médias) e <i>IoU</i> (Interseção sobre União) Médio no conjunto de dados de validação.	42
Figura 4.8 – Comparação dos tempos de resposta para diferentes esquemas de comunicação com o sistema de reconhecimento de objetos.	42
Figura A.1 – Blue Transparent Pen - Caneta Azul Transparente	49
Figura A.2 – Black Transparent Pen - Caneta Preta Transparente	50
Figura A.3 – Black Marker - Marcador Preto	50
Figura A.4 – White Marker - Marcador Branco	51
Figura A.5 – Soccer Ball - Bola de Futebol	51
Figura A.6 – Ping Pong Ball - Bolinha de Ping Pong	52
Figura A.7 – Basketball - Bola de Basquete	52
Figura A.8 – Plain Orange Ball - Bola Lisa Laranja	53
Figura A.9 – Fiction Book - Livro de Ficção	53
Figura A.10 – Journal - Diário	54
Figura A.11 – Movie Encyclopedia - Enciclopédia de Filmes	54
Figura A.12 – Batman Notebook - Caderno do Batman	55

LISTA DE QUADROS

Quadro 3.1 – Especificações da <i>Asus Tinker Board S</i>	19
Quadro 3.2 – Especificações dos aspectos robóticos de Beo	19
Quadro 3.3 – Objetos e algumas características utilizados no cenário de interação. . .	22
Quadro 3.4 – Principais especificações da <i>NVIDIA Jetson TX2</i>	23
Quadro 4.1 – Taxa de detecções falsas por minuto para cada modelo de reconheci- mento de voz em ambiente silencioso e barulhento.	36
Quadro 4.2 – Comparação dos tempos de resposta nas etapas do reconhecimento de fala.	36

SUMÁRIO

1	INTRODUÇÃO	8
1.1	MOTIVAÇÃO	8
1.2	OBJETIVOS	9
1.2.1	Objetivo Geral	9
1.2.2	Objetivos Específicos	9
2	REVISÃO BIBLIOGRÁFICA	10
2.1	INTERAÇÕES HUMANO-ROBÔ	10
2.1.1	Métricas e Validação em Interações Humano-Robô	12
2.2	<i>GAMIFICATION</i>	12
2.2.1	<i>Gamification</i> para Educação	13
2.3	TRABALHOS RELACIONADOS	15
2.3.1	Robôs como Tutores de Língua Estrangeira	15
3	METODOLOGIA	18
3.1	ROBÔ UTILIZADO	18
3.1.1	Quanto ao seu <i>Hardware</i>	19
3.1.2	Quanto à seu <i>Software</i>	20
3.2	OBJETOS FÍSICOS REAIS	21
3.3	NVIDIA JETSON TX2	22
3.4	DESENVOLVIMENTO DO CENÁRIO DE INTERAÇÃO	23
3.4.1	Cenário de Interação: Jogo "Is it?"	24
3.4.1.1	<i>Preparo do Jogo</i>	24
3.4.1.2	<i>Fase Introdução</i>	25
3.4.1.3	<i>Fase Beo Adivinhador</i>	25
3.4.1.4	<i>Fase Usuário Adivinhador</i>	25
3.5	DESENVOLVIMENTO DO SISTEMA DE RECONHECIMENTO DE FALA	26
3.5.1	Testes para medir o desempenho do sistema de reconhecimento de fala	27
3.6	DESENVOLVIMENTO DO SISTEMA DE RECONHECIMENTO DE OBJETOS	28
3.6.1	Métricas e Testes para medir o desempenho	29
4	RESULTADOS	31
4.1	CENÁRIO	31
4.1.1	Transcrição de um cenário	31
4.1.2	Discussão das métricas de IHR e Escalas de Bartneck	34
4.2	RESULTADOS PARA SISTEMA DE RECONHECIMENTO DE FALA	35
4.2.1	Discussão dos resultados para o sistema de reconhecimento de fala	39
4.3	RESULTADOS PARA O SISTEMA DE RECONHECIMENTO DE OBJETOS	39
4.3.1	Discussão de resultados para o sistema de reconhecimento de objetos	43
5	CONSIDERAÇÕES FINAIS	44
	REFERÊNCIAS BIBLIOGRÁFICAS	46
	APÊNDICE A – LISTA COMPLETA DOS OBJETOS FÍSICOS	49
	APÊNDICE B – QUESTIONÁRIO DE ANÁLISE QUALITATIVA DO CENÁRIO DE INTERAÇÃO	57
	APÊNDICE C – TRANSCRIÇÃO DE INTERAÇÃO EM PORTUGUÊS	60

1 INTRODUÇÃO

Neste capítulo é apresentada a motivação que guiou o desenvolvimento deste trabalho e a definição do objetivo geral e dos objetivos específicos que buscam-se cumprir.

1.1 MOTIVAÇÃO

A proficiência em uma segunda língua além da língua nativa é uma habilidade cada vez mais procurada e importante para o desenvolvimento pessoal no mundo globalizado atual. A habilidade de expressão em uma segunda língua permite um indivíduo experimentar culturas além de seu país de origem, proporcionando novas perspectivas e também abrindo oportunidades para o desenvolvimento profissional.

Atualmente, inglês é a língua mais utilizada no mundo, servindo como língua padrão para comunicações internacionais. A maior parte do conhecimento adquirido pela humanidade pode ser facilmente acessado através da *internet*. Porém, para usufruir de tal conhecimento, é necessário que o indivíduo tenha alguma proficiência mínima na língua.

O desenvolvimento cognitivo e aprendizado da linguagem estão intimamente relacionados (SCHUMANN, 1994), e seu aprendizado é extremamente influenciado pela interação social com humanos (KUHL, 2004; KUHL, 2007). Assim, a procura por ferramentas eficazes no ensino de uma segunda língua cresce cada vez mais.

Não é recente o debate sobre mecanismos de motivação para estudantes em sua jornada de aprendizado (BLOOM, 1984), e como evitar a desistência e evasão dos mesmos (BRIDGELAND; JR; MORISON, 2006). Há alguns anos, um grande interesse surgiu no uso de *gamification*, o uso de elementos de *design* de jogos fora do contexto de jogos (DETERDING et al., 2011). De fato, o processo é um bom motivador extrínseco (DETERDING, 2012) e tem seu uso mais comum no contexto da educação (HAMARI; KOIVISTO; SARSA, 2014).

Além disso, com o avanço da tecnologia, robôs modernos se tornaram capazes de conduzir experiências mais interativas, fazendo uso de variados mecanismos de interação social, e, portanto, são considerados uma ferramenta potencialmente eficaz no ensino de uma segunda língua. Robôs fisicamente presentes comprovadamente produzem ganhos significativos no aprendizado (LEYZBERG et al., 2012), e crianças são capazes de aprender elementos de outra língua interagindo com um robô tutor (KENNEDY et al., 2016). Mesmo assim, o campo de interação humano-robô (IHR) está continuamente em busca de mecanismos e técnicas que permitam conduzir experiências cada vez mais naturais para usuários humanos.

1.2 OBJETIVOS

Robôs mostram-se como uma ferramenta com grande potencial no ensino de uma segunda língua, tanto como um elemento auxiliar em sala de aula como um tutor individual. O mesmo potencial pode ser observado no processo de *gamification* amplamente aplicado com sucesso em diversas áreas, inclusive na educação e ensino de línguas estrangeiras. Este trabalho tem inspiração nos conceitos de *gamification* e, baseando-se em estudos de IHR, apresenta uma interação lúdica entre um usuário humano e um robô.

1.2.1 Objetivo Geral

Desenvolvimento de uma plataforma para ensino de língua inglesa, integrando um robô fisicamente presente e sistemas de reconhecimento de fala e objeto.

1.2.2 Objetivos Específicos

- Desenvolver um sistema de reconhecimento de voz.
- Desenvolver um sistema de reconhecimento de objetos.
- Ampliar as capacidades de interação do robô **Beo**, utilizado neste trabalho, segundo diretrizes de IHR.
- Desenvolver um cenário de interação lúdico com objetos físicos e conversação com o robô **Beo**.
- Testar e analisar os sistemas de reconhecimento de fala e de objeto.

2 REVISÃO BIBLIOGRÁFICA

Neste capítulo será apresentada uma revisão de estudos que fundamentam o uso dos elementos técnicos desenvolvidos nesse trabalho, no contexto da interação humano-robô (IHR) e do processo de *gamification*, bem como justificam o próprio uso dos mesmos em um cenário de aprendizado de língua estrangeira.

2.1 INTERAÇÕES HUMANO-ROBÔ

Interação humano-robô é um campo interdisciplinar que estuda interações entre humanos e robôs. Com contribuições de áreas tecnológicas (*i.e.* robótica, inteligência artificial, entre outras) e áreas de estudos sociais e humanas (*i.e.* psicologia, *design*, neurociência, entre outras) busca estabelecer métodos eficazes para o desenvolvimento das interações.

Para que uma interação humano-robô seja eficaz, não basta apenas elementos de eletrônica e *software* avançados, é necessário também um conhecimento de teoria da comunicação e do comportamento humano, moldando as interações de forma a influenciar ou impactar os usuários da maneira desejada. Esforços foram realizados na tentativa de determinar estruturas (*frameworks*) e padrões de *design* que resultam em interações com um alto engajamento humano – eficazes.

Bartneck e Okada (2001) definiram um esquema de classificação com quatro escalas, baseado na aparência e nas capacidades do robô:

- Ferramenta – Brinquedo: A capacidade do robô em auxiliar humanos a realizar alguma tarefa específica ou servir apenas como entretenimento.
- Controle Remoto – Autonomia: A capacidade do robô em operar de forma autônoma ou ser controlado remotamente.
- Reativo – Ativo: A capacidade do robô em tomar a iniciativa em uma interação com humanos ou apenas reagir à ações feitas pelo usuário.
- Antropomorfismo: Relativo à aparência do robô, o quão próximo ele se assemelha da aparência de um humano real.

Bartneck e Okada (2001) também notam algumas observações pertinentes da percepção humana durante interações com robôs. Robôs inúteis, isto é, que não são capazes de ajudar humanos a realizar tarefa alguma, são percebidos como irritantes pelos usuários. Além disso, pessoas atribuem características de personalidades que não foram explicitamente programadas ou pretendidas pelos desenvolvedores do robô.

Kahn et al. (2008) discutiram sobre como padrões de *design* podem ser utilizados para promover o desenvolvimento do campo de IHR. Baseando-se no trabalho do arquiteto Christopher Alexander e sua “linguagem de padrões” (ALEXANDER, 1977), os autores argumentam que o teste e uso de padrões bem-sucedidos permite a criação de uma base fundamental para interações humano-robôs. Através de experimentações é possível definir e descobrir novos padrões, de forma a auxiliar novas pesquisas da área e poupar tempo através do reuso dos padrões mais eficientes. Kahn et al. (2008) propõem métodos para gerar novos padrões de *design* para IHR baseados em quatro diretrizes principais retiradas dos padrões de Alexander:

- Padrões devem ser especificados de forma abstrata o suficiente para que várias instanciações diferentes do padrão possam ser implementadas como solução para um problema.
- Padrões podem e devem ser combinados frequentemente.
- Padrões menos complexos são comumente integrados de forma hierárquica em padrões mais complexos.
- Padrões de *design* são fundamentalmente padrões de interações de humanos com o mundo físico e social.

Através dos padrões busca-se uma qualidade subjetiva, que não é quantificada de forma fácil. Christopher Alexander buscava uma “qualidade sem um nome”, argumentando que pessoas naturalmente preferem padrões melhores. Assim, sem uma fórmula ou um molde rígido, Kahn et al. (2008) propuseram padrões para estimular socialidade em IHR, testados através de experimentações com crianças interagindo com robôs em diversos cenários.

Utilizando desses quatro métodos centrais, os autores propuseram oito padrões de *design* em potencial para IHR. Os padrões foram nomeados de acordo com as oportunidades de interação que eles produzem: *Introdução Inicial*, *Comunicação Didática*, *Movimentando-se Juntos*, *Interesses e História Pessoais*, *Recuperando-se de Erros*, *Tomando Turnos em um Contexto de Jogo*, *Intimidade Física* e *Alegando Tratamento Injusto*. Dentre esses, o autor deste trabalho deseja ressaltar dois padrões fundamentais para o desenvolvimento do trabalho presente:

- *Comunicação Didática*: Uma das formas menos complicadas de comunicação social, a transmissão de informação de um agente à outro, de forma unidirecional. Requerendo pouca responsividade do usuário humano, esse padrão pode ser facilmente implementado com as tecnologias de hoje.
- *Tomando Turnos em um Contexto de Jogo*: Grande parte dos jogos sociais envolve alguma maneira de tomar turnos. A utilização desse padrão é uma maneira eficaz

de manter o engajamento com um usuário humano. Ainda, é possível conectar esse padrão diretamente com o conceito de *gamification*, unindo jogos e IHR.

2.1.1 Métricas e Validação em Interações Humano-Robô

A utilização de mecanismos de IHR está disseminada em uma vasta gama de aplicações diferentes. Essa diversidade, juntamente com o fato de IHR comumente serem um meio para atingir um determinado fim, fez com que grande parte das métricas desenvolvidas sejam específicas à alguma aplicação, de difícil generalização para outras tarefas.

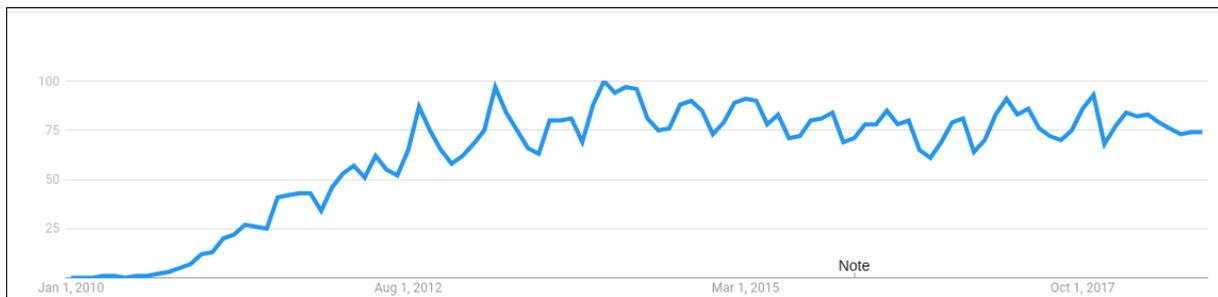
Um dos primeiros esforços para definir princípios de interações eficientes, baseados em métricas quantificáveis foi realizado por Goodrich e Jr (2003). As métricas das quais Goodrich e Jr (2003) utilizaram para definir seus princípios foram baseadas em um robô que necessitava realizar uma tarefa em conjunto com instruções de um humano e refletem questões essenciais no desenvolvimento de IHR. Os sete princípios serão utilizados para moldar as IHR deste trabalho e são apresentados a seguir:

- Implicitamente alternar modos de autonomia e interfaces.
- Permitir que o robô faça uso de sinais humanos naturais.
- Manipular o mundo, ao invés do robô.
- Manipular a relação entre o robô e o mundo.
- Permitir que os usuários manipulem as informações apresentadas.
- Externalizar memória.
- Ajudar os usuários a controlar a atenção.

2.2 GAMIFICATION

Desde o surgimento do termo *gamification* em meados de 2008 e sua popularização em 2010 (DETERDING et al., 2011) pode-se notar um interesse crescente como observado na gráfico da Figura 2.1. Ao longo do tempo, diversos esforços foram realizados para buscar uma definição geral de um processo tão amplamente aplicado em áreas tão variadas (GROH, 2012; WERBACH, 2014; HUOTARI; HAMARI, 2012). Uma definição simples, que abrange diversas áreas de aplicação foi proposta por Deterding et al. (2011) como "o uso de elementos de *design* de jogos fora do contexto de jogos".

Figura 2.1 – Interesse no termo *gamification* desde 2010.



Fonte: Adaptado de *Google Trends*.

Um fator que fortemente suporta o uso de *gamification* em qualquer atividade é sua influência na motivação (DETERDING, 2012; GLOVER, 2013) do usuário. O campo de ensino e educação, que possui grandes problemas em manter estudantes motivados a aprender e evitar evasão (BRIDGELAND; JR; MORISON, 2006), não deixou de demonstrar interesse por esse processo emergente. De fato, o uso de *gamification* na educação é bem documentado (LEE; HAMMER, 2011; DICHEVA et al., 2015) e segundo uma revisão bibliográfica feita por Hamari, Koivisto e Sarsa (2014) esse é o contexto de uso mais comum.

2.2.1 *Gamification* para Educação

No cenário atual há uma vasta gama de produtos que utilizam os conceitos de *gamification* para o ensino de segundas línguas, de aplicativos *mobile* como **Duolingo** (DUOLINGO, 2018) a *softwares* completos para *desktop* com modelos de assinatura como **Rosetta Stone** (STONE, 2018). Como Flores (2015) discute em sua investigação, o uso de *gamification* contribui positivamente no aprendizado de língua estrangeira, criando experiências significativas que melhoram o aprendizado de escrita, leitura, fala e motivam a colaboração e interação do aluno.

Atividades educacionais que façam uso de *gamification* devem ser interativas, engajadoras e ricas em elementos multimídia, e o desenvolvimento das mesmas deve possuir foco nos objetivos de aprendizado (KIRYAKOVA; ANGELOVA; YORDANOVA, 2014). Simões, Redondo e Vilas (2013) propuseram diretrizes para o desenvolvimento de ferramentas de ensino e *gamification*, com o intuito de providenciar ferramentas de ensino engajadoras e melhorar a motivação e aprendizado dos estudantes. Dentre as diretrizes propostas, no que se diz respeito ao desenvolvimento de atividades por meio de *gamification*, elenca-se as seguintes:

- **Desempenho Variado:** As atividades devem ser desenvolvidas de modo que os usuários possam repeti-las em caso de uma tentativa falha.

- **Viabilidade:** As atividades devem ser possíveis de realizar. A dificuldade e complexidade deve ser adaptada para o nível do usuário e suas habilidades.
- **Aumento da Dificuldade:** É esperado que cada atividade subsequente seja mais complexa que a anterior, requerendo mais esforço do usuário e correspondendo às suas habilidades e conhecimento recém adquiridas.
- **Diversos Caminhos:** Para desenvolver habilidades variadas em usuários, eles precisam ser capazes de completar os objetivos de maneiras variadas. Isso permite que usuários construam suas próprias estratégias, uma das características principais no aprendizado ativo.

O principal objetivo de *gamification* no ensino é aumentar o engajamento e participação dos usuários, motivando-os através do uso de elementos de *design* de jogos, como pontos, recompensas e respostas imediatas (FLORES, 2015). O debate sobre o uso de tecnologia e computadores no auxílio do ensino de língua estrangeiras é de longa data e precede a popularização do termo *gamification* (YBARRA; GREEN, 2003; HUBBARD; LEVY, 2006). Atualmente existem inúmeras aplicações que fazem uso de *gamification*, onde grande parte delas consistem em aplicativos *Web* e *mobile* para plataformas *Android* e *iOS*. De fato, o processo atingiu tamanha popularização que atualmente pode ser difícil encontrar um aplicativo voltado para educação que não faz uso de algum elemento de *gamification*. Alguns exemplos populares de aplicativos para o ensino de língua estrangeira são:

- **Duolingo:** Pode ser acessada pela *Web* e também possui aplicativo para *Android* e *iOS*. Atualmente conta com uma base de aproximadamente 300 milhões de usuários e suporte para 37 línguas. As lições são divididas em níveis e usuários possuem fortes estímulos para o uso diário do aplicativo, através do uso de elementos como sistemas de recompensa, experiência e uma moeda própria que os usuários podem utilizar para comprar itens virtuais dentro do aplicativo. (DUOLINGO, 2018)
- **ClassDojo:** Plataforma permite que estudantes, pais e professores se conectem para fornecer suporte e ajudar a desenvolver as habilidades do estudante. Acessível pela *Web* ou sistemas *Android* e *iOS*. Professores podem se registrar e criar aulas com os estudantes, postando fotos e vídeos e acompanhando o progresso dos mesmos. Os estudantes podem customizar personagens virtuais com recompensas fornecidas pelos professores por bom desempenho. (CLASSDOJO, 2018)
- **MindSnacks:** Aplicativo para *iOS*, com módulo para ensino de línguas. Ensina vocabulário, escrita e gramática através de mini jogos. Permite o acompanhamento do progresso do usuário através de pontos e experiência recompensada por bom desempenho. (MINDSNACKS, 2018)

Gamification é incorporado no desenvolvimento de aplicativos e plataformas *Web* com comprovado sucesso (VESSELINOV; GREGO, 2012; OSIPOV et al., 2015). Como previamente discutido, o aprendizado de línguas está intimamente ligado com interações sociais (KUHL, 2004). Algumas plataformas como **Edmodo** (EDMODO, 2018) foram desenvolvidas seguindo modelos de redes sociais, como o *Facebook*, incentivando o aprendizado através da interação social. Também existem aplicativos como **HelloTalk** (HELLO-TALK, 2018) que conectam usuários de diferentes países para conversar e trocar conhecimento sobre suas línguas nativas. Ainda assim, alguns indivíduos se sentem constrangidos ao interagir com outras pessoas sem um certo domínio da língua. Robôs podem ser um substituto potencial para falantes iniciantes, promovendo um ambiente mais confortável, sem julgamento. Levando em consideração que a simples presença física de um robô resulta em melhor desempenho cognitivo quando comparado com uma interação com um avatar virtual (LEYZBERG et al., 2012), a aplicação de *gamification* no desenvolvimento de robôs tutores de língua estrangeira é um contexto interessante ainda a ser amplamente explorado.

2.3 TRABALHOS RELACIONADOS

Nesta seção são apresentados trabalhos relacionados que utilizam robôs e processos de *gamification* no ensino de língua estrangeira.

2.3.1 Robôs como Tutores de Língua Estrangeira

Uma aula comum é tradicionalmente lecionada em horários pré-estabelecidos com um professor servindo de tutor para um grupo de alunos. Essa tradição nem sempre leva a bons resultados no aprendizado, seja por indisposição dos alunos e professores, ou até mesmo falta de atenção individual ao aluno. De fato, a busca por métodos de ensino tão eficazes quanto aulas particulares não é recente (BLOOM, 1984; DELQUADRI et al., 1986). A utilização de robôs como ferramentas de ensino é um campo de pesquisa emergente da área de IHR com potencial para solucionar esses problemas.

Kanda et al. (2004) realizaram um experimento com crianças de 6 a 12 anos de idade. Dois robôs **Robovie** (Figura 2.2) foram colocados nos perímetros das salas de aula e as crianças foram deixadas livres para interagir com os mesmos. Com capacidade de reconhecimento de aproximadamente 50 palavras e 300 frases pré-gravadas em inglês, o objetivo era examinar se crianças conseguem formar um relacionamento com um robô e aprender com eles como elas aprendem com outras crianças. Os resultados demonstram que as crianças construíram uma relação com os robôs, levando algumas delas a melhorar

seu inglês naturalmente com o progresso das interações. Porém, o estudo também serve para apontar algumas das limitações de robôs em situações sociais reais. Após a primeira semana, o número de interações com os robôs diminuiu significativamente, sugerindo que após um período de novidade as crianças perderam o interesse. Isso destaca a importância de desenvolver esquemas e padrões robustos de interações que consigam atingir um alto engajamento, possivelmente à longo prazo (BARTNECK; OKADA, 2001; KAHN et al., 2008).

Figura 2.2 – **Robovie** de Kanda et al. (2004) interagindo com uma criança.

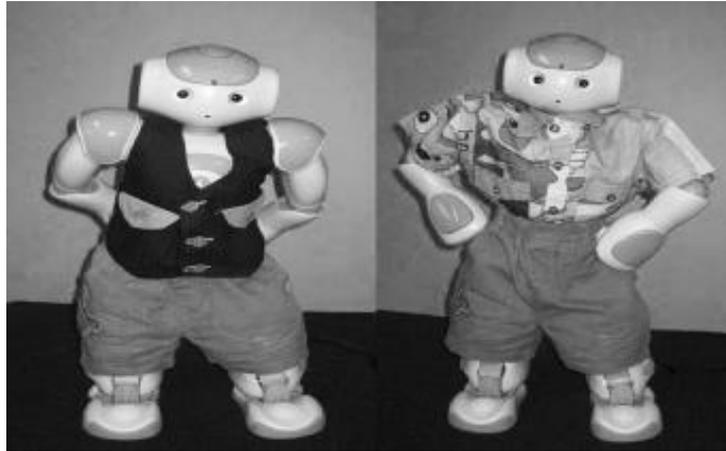


Fonte: Adaptado de Kanda et al. (2004)

Em um trabalho mais recente, Alemi, Meghdari e Ghazisaedy (2014) utilizaram o robô **NIMA** (baseado na plataforma **NAO**) (Figura 2.3) em sala de aula, juntamente com um tutor humano para ensinar a língua inglesa para estudantes entre 12 e 15 anos de idade. O robô foi programado para explicar palavras do vocabulário inglês através de ações e sons, utilizando-as em diálogos simples com o professor. O objetivo da pesquisa era investigar o aprendizado de vocabulário através de uma interação conjunta com o tutor e o robô. Para tal, os estudantes foram divididos em dois grupos: um que participou de interações com o robô e outro que não interagiu com o robô. Os resultados demonstram que o grupo que interagiu com o robô atingiu uma média significativamente maior que o grupo que não interagiu com o robô, tanto em um teste imediatamente após o término do experimento e um teste de retenção duas semanas mais tarde. Esse trabalho apresenta evidências fortes que o uso de robôs pode ser uma ferramenta útil no aprendizado e na retenção do vocabulário de línguas estrangeiras quando utilizado em conjunto com um tutor humano.

Um fator bastante discutido em pesquisas relacionadas à educação e aprendizado é a questão da disponibilidade do tutor (WITT; WHEELLESS; ALLEN, 2004; GORHAM, 1988). Utilizando um robô em um contexto com atenção individual, Kennedy et al. (2016) estudaram o efeito empregando um **NAO** para ensinar francês para 67 crianças de 8 e 9

Figura 2.3 – Robô **NIMA**, utilizado por Alemi, Meghdari e Ghazisaedy (2014) para ensinar crianças a língua inglesa através de estórias.



Fonte: Adaptado de Alemi, Meghdari e Ghazisaedy (2014).

anos de idade. Kennedy et al. (2016) definiram dois estados de disponibilidade verbal: alta, onde o robô engajava ativamente com a criança usando o seu nome, perguntando sobre seus interesses e *hobbies*, etc.; e baixa, onde o robô apenas respondia de forma curta e simples. Os estudantes interagiram com uma *touchscreen* respondendo questionários de língua francesa enquanto o robô assumia o papel de um tutor tradicional, semelhante ao padrão *Comunicação Didática* (KAHN et al., 2008). Os resultados demonstram que as crianças que interagiram com robô obtiveram resultados melhores nos testes de vocabulário logo após o experimento e outro teste de retenção, uma semana após, quando comparados à um grupo de controle, que não interagiu com o robô. Interessantemente, os resultados apresentaram pouca variação entre os grupos de crianças que interagiram com o robô de alta e o de baixa disponibilidade verbal, contrariando algumas pesquisas prévias da área (KRUIJFF-KORBAYOVA et al., 2014; HENKEMANS et al., 2013). Apesar da variação entre alta e baixa disponibilidade verbal não ter resultado em diferenças de melhora no aprendizado, isso não descarta o uso dos robôs como ferramenta de ensino, tanto individual quanto para um grupo. De fato, uma possível explicação para esse fenômeno é que a simples presença física de um robô melhora os ganhos cognitivos dos alunos (LEYZBERG et al., 2012), e um investimento maior em uma alta disponibilidade verbal, pode não resultar em um retorno cognitivo proporcional (KENNEDY et al., 2016).

3 METODOLOGIA

Nesse capítulo serão apresentados e detalhados os passos de variados aspectos do desenvolvimento deste trabalho. Nas Seções 3.1, 3.2 e 3.3 serão apresentados os materiais utilizados, incluindo o robô, seu *software* e *hardware*, os objetos físicos e *hardwares* externos. Na Seção 3.4 é apresentado o desenvolvimento do cenário de jogo. Nas Seções 3.5 e 3.6 serão detalhados o desenvolvimento dos sistemas de reconhecimento de fala e objetos, bem como testes relevantes efetuados.

3.1 ROBÔ UTILIZADO

O **Beo** (Figura 3.1) é um robô desenvolvido pela *QironRobotics*¹, empresa que desenvolve robôs de código aberto e fornece cursos de robótica para crianças e adultos. Com um *design* amigável e carcassa impressa em 3D, **Beo** é utilizado em ambientes de sala de aula, auxiliando o ensino da programação e interagindo de maneira simplística com roteiros pré-programados.

Figura 3.1 – Robô **Beo**.



Fonte: Disponível em <<http://qironrobotics.com/robos/>>.

¹<<http://qironrobotics.com>>

3.1.1 Quanto ao seu *Hardware*

Beo conta como sua principal unidade de processamento um SBC (Em inglês, *Single Board Computer* – Computador de Placa Única). Esse tipo de computador é comumente utilizado na construção de robôs por reunir todos os componentes necessários para um computador funcional em uma única placa (Processador, Entrada e Saída, Armazenamento, Memória, etc.). Originalmente, **Beo** utilizava um *Raspberry Pi 3*, todavia um *upgrade* para uma *Asus Tinker Board S* (Figura 3.2) foi realizado para melhor suportar os algoritmos e métodos desenvolvidos para este trabalho. As especificações do computador "coração" de **Beo** são mostradas no Quadro 3.1.

Quadro 3.1 – Especificações da *Asus Tinker Board S*.

CPU	Quad-core 1.8 GHz ARM Cortex-A17
GPU	600 MHz Mali-T760
Memória RAM	2 GB LPDDR3
Armazenamento	16 GB eMMC interno
Dimensões	8.55 cm x 5.4 cm
Peso	55g

Fonte: Adaptado de <<https://www.asus.com/us/Single-Board-Computer/Tinker-Board-S/specifications/>>.

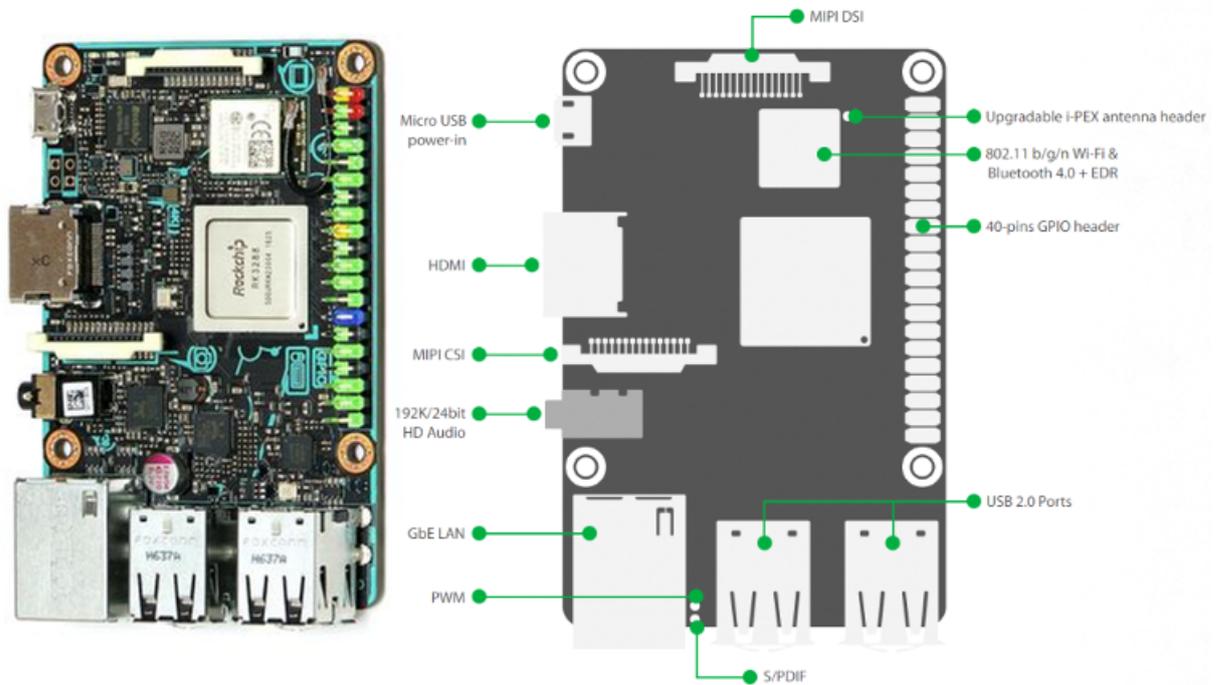
Convém lembrar que o desempenho de um robô em realizar determinadas tarefas não depende apenas de seu recursos computacionais, mas também de seus recursos "físicos", a capacidade do robô interagir com o mundo real. Além do mais, no que tange interações humano-robô, o aspecto físico ou seja, a aparência do robô, afeta a percepção do usuário sobre toda a experiência de interação. No Quadro 3.2 são apresentados detalhes técnicos que abrangem alguns desses aspectos para o robô **Beo**.

Quadro 3.2 – Especificações dos aspectos robóticos de **Beo**.

Carcaça	Impresso 3D (PLA)
Computador	ASUS Tinker Board S
Câmera	Logitech C270
Motores	Dynamixel XL-320
Bateria	LiFE 3s 2200 mAh - Autonomia de 2h
Sensores de Navegação	Encoder de Quadratura
Movimento	Rodas 4"
Graus de Liberdade	9 (3 para cabeça e 3 para cada braço)
Caixa de Som	15W
Microfone	Sim
Variados	Sensor de toque capacitivo, Display OLED para olhos

Fonte: Adaptado de <<http://qironrobotics.com/robos/>>.

Figura 3.2 – Foto de Tinker Board (à esquerda) e esquemático de entradas e saídas (à direita).



Fonte: Adaptado de <<https://www.asus.com/br/Single-Board-Computer/Tinker-Board/>>

3.1.2 Quanto à seu *Software*

Beo atualmente é um robô limitado no que consta suas capacidades de IHR de forma autônoma. Todavia, **Beo** possui um *hardware* perfeitamente capaz de suportar os algoritmos e métodos mais complexos necessários para interações IHR. De fato, levando em consideração que **Beo** já possui um *design* amigável, fator de grande importância, que é comprovadamente capaz de engajar estudantes, ele é uma plataforma com um grande potencial para o ensino de uma língua estrangeira, objetivo desse trabalho. Assim, grande parte dos esforços foram direcionados à estender e dar novas capacidades ao **Beo**, além das interações básicas, promovendo uma experiência mais completa para o usuário.

Seguindo na mesma veia que as aplicações de robótica da atualidade, **Beo** roda um sistema operacional baseado em *Linux*, *TinkerOS 2.0* distribuído pela própria fabricante *Asus*. A ampla disseminação de sistemas baseados em *Linux* na área é justificada pelo fato de ser um sistema operacional de código aberto, leve (na maior parte de suas distribuições) e altamente customizável.

Também de uso muito comum na área da robótica é o **ROS**. **ROS** é um meta-sistema operacional, um esquema pacotes para comunicação entre módulos com funcionalidades diferentes. Cada versão de **ROS** é desenvolvida visando suportar distribuições específicas de *Linux*. Assim, considerando que *TinkerOS 2.0* é um sistema baseado em *Debian 9*, a versão de **ROS** adequada e utilizada foi a *Lunar*. Para o desenvolvimento

desse trabalho, primeiramente foram reimplementados todos os módulos básicos do **Beo** em pacotes **ROS**. Esses módulos incluem códigos que realizam comunicação com os motores, reprodução de áudio, fala, controle de movimentos, etc. O uso de **ROS** permite que os códigos sejam escritos tanto em *C++* ou *Python*.

3.2 OBJETOS FÍSICOS REAIS

O cenário de interação consiste em um jogo de adivinhação em turnos jogado entre robô e o usuário humano. Nesse jogo, objetos do mundo real são utilizados (Figura 3.3, com a intenção de ligar palavras da língua inglesa com objetos que podem ser fisicamente manipulados. De maneira a não tornar o jogo de adivinhação fácil demais, os objetos propostos são variações de objetos pertencentes a uma mesma categoria: canetas, bolas e cadernos. Assim, os objetos possuem algumas características em comum, porém ainda mantendo a adivinhação correta do objeto específico um tanto desafiadora. O Quadro 3.3 brevemente apresenta as categorias propostas juntamente com os objetos escolhidos e suas características em inglês, utilizados durante o jogo. Uma lista completa de todas as características e fotografias para cada objeto utilizado neste trabalho é apresentado no Apêndice A.

Outro tipos de artifício com implementação mais prática poderiam ter sido utilizados nesses cenários de *minigames*, como por exemplo um conjunto de cartas com figuras de objetos. A escolha por objetos físicos é justificada por apresentarem uma ligação direta com a palavra em inglês durante o aprendizado da língua, sem a necessidade de uma tradução através de um símbolo intermediário. Além disso, espera-se que o manuseio de objetos presentes resulte em uma interação mais divertida para o usuário.

Figura 3.3 – Fotografia dos objetos físicos espalhados pelo chão.



Fonte: Autor.

Quadro 3.3 – Objetos e algumas características utilizados no cenário de interação.

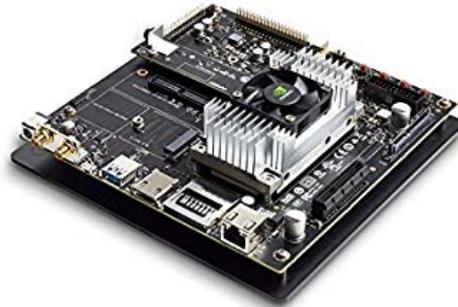
Categoria: Balls (Bolas)	Características
Plain Orange Ball	plain, orange, squishy, soft...
Soccer Ball	hard, striped, blue, white, light...
Ping Pong Ball	tiny, hard, bouncy, plain...
Basketball	large, bouncy, blue, heavy...
Categoria: Books (Livros e Cadernos)	
Fiction Book	white, yellow, medium-sized, fiction...
Journal	brown, thin, blank, soft-cover, notebook...
Batman Notebook	colorful, rectangular, hard-cover, light...
Movie Encyclopedia	colorful, thick, heavy, movies, large, big...
Categoria: Pens (Canetas)	
White Marker	white, black-tipped, light, cylinder, thick...
Black Marker	black, blue, cylinder, light, thick...
Blue Transparent Pen	transparent, blue, ink, light...
Black Transparent Pen	transparent, black, ink, light, rubbery...

Fonte: Autor.

3.3 NVIDIA JETSON TX2

A *NVIDIA Jetson TX2* é uma placa desenvolvida pela NVIDIA contando com uma GPU com grande poder de processamento e baixa potência, com o intuito de executar modelos de redes neurais e visão computacional em tempo real. Por ser o melhor *hardware* acessível durante o desenvolvimento deste trabalho, a *Jetson TX2* foi utilizada para treinar uma rede neural para reconhecimento de objetos físicos. Através do uso de **ROS**, também foi utilizada para executar o modelo treinado, comunicando-se com o robô. Uma descrição completa de suas especificações é apresentada no Quadro 3.4 e uma fotografia da placa

Figura 3.4 – Placa de Desenvolvimento *NVIDIA Jetson TX2*.



Fonte: Adaptado de <<https://www.nvidia.com/en-us/autonomous-machines/embedded-systems-dev-kits-modules/>>.

é mostrada na Figura 3.4.

Quadro 3.4 – Principais especificações da *NVIDIA Jetson TX2*.

GPU	NVIDIA Pascal, 256 Núcleos CUDA
CPU	HMP Dual Denver 2 + Quad-Core ARM A57
Memória RAM	8 GB LPDDR4
Armazenamento	32 GB eMMC interno

Fonte: Adaptado de <<https://www.nvidia.com/en-us/autonomous-machines/embedded-systems-dev-kits-modules/>>.

3.4 DESENVOLVIMENTO DO CENÁRIO DE INTERAÇÃO

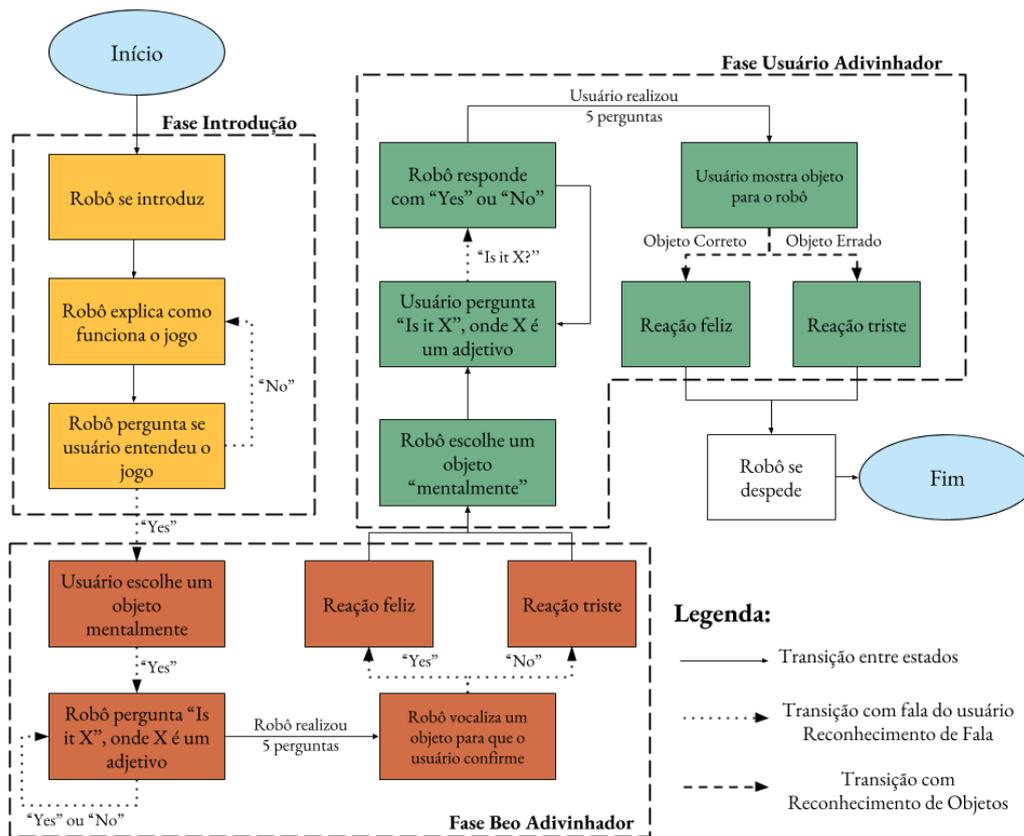
O objetivo do cenário de interação é o ensino da língua inglesa através de um jogo de interação humano-robô. O cenário foi desenvolvido visando proporcionar um ambiente de baixo estresse e divertido, reforçando o uso frequente da atividade com o robô.

O cenário de interação faz uso de objetos físicos juntamente com elementos de conversação com o robô para ensinar palavras e adjetivos da língua inglesa. Foram escolhidos 12 objetos físicos reais e de uso comum, com o critério de que não fossem grandes demais, tornando difícil de armazená-los e também não fossem pequenos demais, dificultando a detecção dos mesmos pela câmera do robô. Feita a escolha dos 12 objetos eles foram rotulados de acordo com as características físicas que apresentam, tangendo tamanho, aparência, formato, peso, etc.

3.4.1 Cenário de Interação: Jogo "Is it?"

O usuário e robô revezam-se selecionando um objeto enquanto o outro jogador deve adivinhar o objeto escolhido através de perguntas em uma interação de conversação. O objetivo deste cenário é o reforço do aprendizado de vocabulário através de um jogo simples e engajante. Nesse jogo o usuário deve falar, ouvir e compreender frases simples com formatos pré-definidos. O fluxo geral do jogo e suas fases pode ser observado no fluxograma da Figura 3.5.

Figura 3.5 – Fluxo do jogo de interação com o Beo.



Fonte: Autor.

3.4.1.1 Preparo do Jogo

O usuário e o robô se posicionam frente à frente com os 12 objetos físicos dispostos no chão entre eles, como demonstrado na Figura 3.6.

Figura 3.6 – Disposição dos jogadores e objetos físicos para o início do jogo.



Fonte: Autor.

3.4.1.2 Fase Introdução

Nessa fase o robô **Beo** se introduz para o usuário e pergunta se ele deseja jogar um jogo. Após, o robô explica as regras do jogo e pergunta se o usuário entendeu. O usuário deve responder com “Yes” ou “No”, detectado pelo sistema de reconhecimento de fala. Caso afirmativo, prossegue-se para a próxima fase, caso contrário o robô explica as regras novamente.

3.4.1.3 Fase Beo Adivinhador

Beo anuncia que irá começar como o adivinhador e, portanto, o usuário deve escolher um dos objetos à sua frente. Quando o usuário tiver selecionado o objeto deve falar “Yes”. O robô então realizará cinco perguntas sobre as características do objeto no formato “Is it X”, onde X é uma característica do objeto. O usuário deve responder às perguntas com “Yes” ou “No”. Após realizar cinco perguntas o robô fará uma tentativa de adivinhar o objeto escolhido pelo usuário, vocalizando o nome de um dos objetos. O usuário, então, deve responder “Yes” caso o robô tenha escolhido o objeto corretamente, gerando uma reação feliz de **Beo**. Caso o robô tenha escolhido incorretamente, o usuário deve responder “No”, reagindo de forma triste. Prossegue-se para a próxima fase.

3.4.1.4 Fase Usuário Adivinhador

Nessa fase os papéis são invertidos. O robô escolhe um objeto e anuncia para o jogador que está pronto para responder as perguntas. O usuário deve fazer cinco perguntas no mesmo formato “Is it X”, que o robô responderá com “Yes” ou “No”. Após as cinco

perguntas o robô requisitará que o usuário mostre um objeto, detectado pelo sistema de reconhecimento de objetos. Caso o usuário tenha mostrado o objeto correto, o robô reage de forma feliz ou infeliz caso tenha mostrado o objeto incorreto. Finalmente, o robô se despede e dá-se fim ao jogo.

3.5 DESENVOLVIMENTO DO SISTEMA DE RECONHECIMENTO DE FALA

O cenário de interação com o robô envolve várias etapas de conversação, algumas em que o usuário responde de forma simples com apenas uma palavra (e.g. "Yes" ou "No"), e outras mais complexas (e.g. "Is it large?", "Is it colorful?"). Dessa maneira, é necessário que o sistema de reconhecimento seja capaz de processar frases de complexidade variada em tempo hábil, para que a conversação entre o usuário e robô proceda de forma natural. O sistema desenvolvido consiste, então, em uma combinação de dois sistemas de reconhecimento bem estabelecidos, porém com funções distintas.

- **Snowboy** (SNOWBOY, 2018): Um sistema leve de reconhecimento de palavras-chave. Utilizando redes neurais e gravações de voz com vocalizações da palavra-chave, gera-se um modelo de detecção que pode ser utilizado *offline*. Em sua execução, o sistema analisa continuamente o som do ambiente, em busca da vocalização da palavra-chave.
- **Google Speech-to-Text** (GOOGLE, 2018): Um serviço na nuvem, que permite que uma gravação de voz seja enviada e retorna as vocalizações em formato de texto. É necessária conexão com a *internet*.

Na Figura 3.7 é possível observar as etapas do processo de reconhecimento da fala completo. As Etapas 1 e 2, demonstradas em azul, são a parte do processo que utiliza *Snowboy* e é realizada totalmente *offline*, o que resulta em um tempo de resposta pequeno. Em contrapartida na Etapa 3, em vermelho, é utilizado o serviço de *Speech-to-Text* do *Google*, com um tempo de resposta maior, devido à necessidade de comunicação com a *internet*.

Foram treinados três modelos para a detecção de três palavras-chave: "Yes", "No" e "Is it". Observando as fases do cenário de interação apresentadas no fluxograma da Figura 3.5, destaca-se que os modelos de "Yes" e "No" são utilizados nas três fases do cenário, enquanto o modelo "Is it" é utilizado somente na fase **Usuário Adivinhador**. O sistema de reconhecimento de fala é utilizado para realizar as transições de estado a partir de uma fala do usuário e possui dois modos de funcionamento distintos:

- 1. Transição de "Yes" e "No": Como o objetivo é apenas determinar se uma das

palavras foi vocalizada, essa transição é realizada na Etapa 1 do sistema de reconhecimento de fala. Utilizado em todas as fases do cenário de interação.

- 2. Transição de “Is it”: Utilizado na fase **Usuário Adivinhador** do cenário de interação. Procede por todas as etapas do sistema de reconhecimento de fala. Na Etapa 1, aguarda-se a vocalização de “Is it”. Na Etapa 2, grava-se as vocalizações posteriores. Na Etapa 3, é enviada a gravação posterior para o serviço de Fala-para-Texto do *Google* e é recebido sua transcrição textual em resposta. Por exemplo, se o usuário vocalizar “Is it round?”, ao final da Etapa 3 o resultado será “round”.

Figura 3.7 – Ordem dos processos do sistema reconhecimento de fala completo.



Fonte: Autor

3.5.1 Testes para medir o desempenho do sistema de reconhecimento de fala

Para determinar a robustez o sistema de reconhecimento de fala, foram realizados testes variados para medir taxas de detecções falsas, precisão e tempos de resposta. A fim de mensurar o desempenho sob diferentes condições os testes 1, 2 e 3 foram realizados em dois ambientes distintos: Silencioso e Barulhento. O ambiente silencioso é uma sala sem presença de pessoas ou equipamentos que produzam sons. O ambiente barulhento é uma área de trabalho com diversas pessoas durante o horário de expediente. Sons no ambiente barulhento incluem pessoas falando, computadores, impressoras, etc.

- **1. Teste de Falsos Positivos:** O teste é realizado com os 3 modelos de palavra-chave utilizados pelo *Snowboy*. O sistema de reconhecimento de fala é executado

por uma hora nos dois ambientes definidos anteriormente, e um contador é incrementado para cada detecção. Nenhuma vocalização de alguma das palavras-chave é realizada durante esse período.

- **2. Teste de Precisão da Etapa 1:** O teste é realizado com os 3 modelos de palavra-chave utilizados pelo *Snowboy*. Para cada modelo, sua palavra-chave é vocalizada 50 vezes e um contador é incrementado para cada detecção. O teste é realizado nos dois ambientes.
- **3. Teste de Precisão da Etapa 3:** O teste é realizado para cada característica possível dos objetos. Para cada característica, a vocalização da palavra é repetida 15 vezes. O teste é realizado nos dois ambientes.
- **4. Teste de tempo de resposta:** Define-se o tempo de resposta como o intervalo de tempo entre o instante em que o usuário finaliza sua vocalização e o instante em que o robô começa sua próxima vocalização. O teste é realizado automatizando a vocalização de cada palavra-chave e medindo o tempo até sua detecção.

3.6 DESENVOLVIMENTO DO SISTEMA DE RECONHECIMENTO DE OBJETOS

Na fase **Usuário Adivinhador** do cenário de interação (Figura 3.5), no penúltimo estado o usuário deve “mostrar” um objeto para o robô. Para realizar a tarefa de reconhecimento de objetos foi treinado um modelo de Rede Neural de Convolução (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). O modelo escolhido foi *Tiny YOLO V3* (REDMON; FARHADI, 2016), por ser um modelo estado da arte que possui foco em detecção em tempo real. O treinamento do modelo ocorreu da seguinte forma:

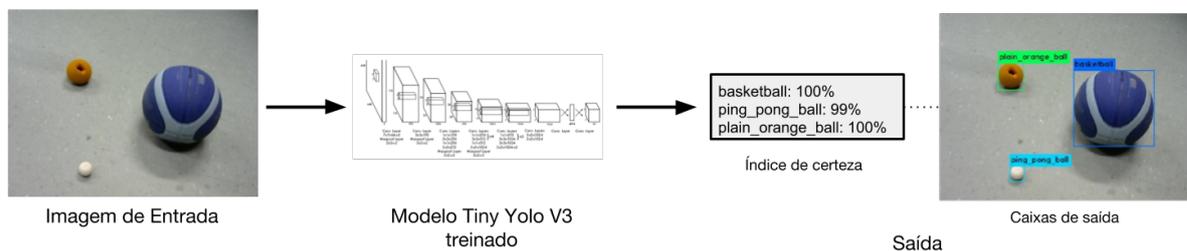
- 1. Para cada objeto isolado foram tiradas aproximadamente 300 fotos com variações de ângulo, luminosidade e presença de outros objetos na vizinhança. Além disso foram tiradas 120 fotos dos objetos em conjunto na mesma imagem, totalizando 3746 imagens e 12 classes de objetos.
- 2. Para cada uma das 3746 imagens foi realizada uma anotação contendo as classes juntamente com caixas que delimitam a posição e tamanho dos objetos presentes na imagem.
- 3. Após a anotação de cada imagem, o conjunto de dados (imagens mais anotações) foi separado em 3458 imagens para treino e 288 para validação. As 288 imagens de validação não serão fornecidas ao modelo durante o treinamento, e servirão apenas para determinar o desempenho do modelo após o mesmo.

- 4. Utilizando a *NVIDIA Jetson TX2* o modelo foi treinado por aproximadamente 72 horas nas 3458 imagens para treino.

Modelos de rede neural são notórios por consumirem muita RAM e processamento, impossibilitando a sua execução pelo *SBC* (Quadro 3.1) do próprio robô. Assim, o sistema de reconhecimento de objetos é executado externamente na *Jetson TX2*. A conexão pode ser realizada de forma cabeada (*Ethernet*) ou *Wi-Fi* e a interface de comunicação entre o *SBC* do robô e a *Jetson TX2* é realizada através de **ROS**.

Após o treinamento, uma imagem pode ser fornecida para o modelo, que retornará uma porcentagem de certeza juntamente com caixas que contém os objetos detectados, como demonstrado na Figura 3.8. Com a finalidade de melhorar a robustez da detecção e diminuir o número de falsos positivos, cada quadro da câmera é enviado para o modelo e um contador específico para cada objeto é incrementado quando sua presença é detectada. Considerando que o modelo *Tiny YOLO V3* opera a uma taxa de aproximadamente 28 quadros por segundo, foi determinado que para uma detecção ser considerada verdadeira é necessária a detecção do mesmo objeto em 25 quadros consecutivos, número em que há um bom balanço entre tempo de resposta e robustez de detecção.

Figura 3.8 – Entrada e saída de uma imagem com o modelo Tiny Yolo V3 treinado.



Fonte: Autor

3.6.1 Métricas e Testes para medir o desempenho

O objetivo da etapa de treinamento da rede neural é minimizar o valor do erro, ajustando um conjunto de parâmetros que controla os valores de saída. Assim, uma métrica que nos permite julgar seu desempenho preliminarmente é o valor do erro durante e ao final do treinamento. Entretanto, o erro é uma parte intrínseca de um modelo e, portanto, não é uma métrica robusta quando se deseja avaliar o desempenho do mesmo de uma forma geral.

Para determinar o desempenho de um modelo de reconhecimento de objetos de forma mais geral e comparável a outros modelos, comumente utiliza-se as métricas *mAP* – *Mean Average Precision* – (do Inglês, Média das Precisoões Médias) e *IoU* – *Intersection over Union* – (do Inglês, Interseção sobre União) (REDMON; FARHADI, 2016).

Para determinar o mAP , primeiramente calcula-se a AP – *Average Precision* – para cada uma das classes do conjunto de dados, i.e. percentagem de predições corretas por classe. mAP é definido, então, como a média das precisões de todas as classes. A Equação (3.1) define mAP matematicamente, em que C é o conjunto de todas as classes, N_C é o número de classes e AP_c é a precisão média por classe.

Para determinar IoU , calcula-se a área da interseção da predição da caixa de saída com a caixa anotada e divide-se pela área da união das mesmas. Matematicamente, a Equação (3.2) define IoU , em que Box_p é a predição do modelo e Box_t é a caixa anotada. De maneira intuitiva, IoU quantifica o quão bem a caixa de predição se “encaixa” na caixa anotada. Quando IoU tem valor 1 significa que a caixa prevista é exatamente igual à caixa anotada.

$$mAP = \frac{1}{N_C} \sum_{c \in C} AP_c \quad (3.1)$$

$$IoU = \frac{area(Box_p \cap Box_t)}{area(Box_p \cup Box_t)} \quad (3.2)$$

O tempo de resposta do sistema de reconhecimento de fala é definido como o intervalo de tempo entre o instante em que a imagem é enviada pelo *SBC* via **ROS** e o instante em que se obtém a resposta com o objeto detectado. O teste é realizado enviando continuamente uma imagem “negativa”, que não contém um objeto detectável pelo sistema. Após um período de tempo envia-se uma imagem positiva, que contém um objeto detectável, e inicia-se um contador de tempo. Quando recebida a mensagem com o objeto detectado, para-se o contador e mede-se o tempo de resposta. O teste foi repetido para três esquemas de conexão distintos: *Cabeado (Ethernet)*, *Wi-Fi 2.4GHz* e *Wi-Fi 5GHz*.

4 RESULTADOS

Nesse capítulo serão apresentados os resultados dos testes dos sistemas de reconhecimento de objetos e voz, além de uma transcrição completa de um cenário de interação entre o robô **Beo** e um usuário.

4.1 CENÁRIO

Na Seção 4.1.1 é apresentada uma transcrição completa do cenário de interação entre o robô e um usuário humano. Com o cenário desenvolvido, uma análise posterior foi realizada pelo autor para julgar a colocação de **Beo** segundo as escalas de Bartneck e Okada (2001), apresentado na Figura 4.1a. Para fins de comparação, a mesma análise foi realizada para os robôs **NIMA** de Alemi, Meghdari e Ghazisaedy (2014) e **Robovie** de Kanda et al. (2004), apresentados nas Figuras 4.1b e 4.1c, respectivamente.

4.1.1 Transcrição de um cenário

Do mesmo modo que o cenário de interação consiste de uma conversação em inglês, a transcrição aqui também é apresentada em inglês. Uma versão traduzida para português pode ser conferida no Apêndice C.

Beo: Hello [acenando]! Wanna play a game [piscadela]?

Beo: This will be a guessing game with these objects in front of us [apontando para os objetos].

Beo: Each of us will take turns in choosing an object, and the other person has to guess what it is.

Beo: We will do that by asking “Yes” and “No” questions about the characteristics of these objects.

Beo: All our questions should start with “Is it?”. For example, “Is it big?”, “Is it white?” or “Is it a cylinder?”.

Beo: Did you understand everything so far?

Usuário: Yes.

Beo: Great! Let's start.

Beo: Okay. I'll start as the guesser. Choose one of the objects in front of me. But don't tell me.

[pausa]

Usuário sinaliza que escolheu o objeto Bola de Basquete.

Beo: When you are ready, say "Yes".

Usuário: Yes.

Beo: Okay. I'm going to try and guess which object you chose.

Beo: Answer my questions!

Beo: Is it rectangular?

Usuário: No.

Beo: Is it pen?

Usuário: No.

Beo: Is it white?

Usuário: No.

Beo: Is it round?

Usuário: Yes.

Beo: Is it bouncy?

Usuário: Yes.

Beo: You're thinking of the basketball, right?

Usuário: Yes.

Beo: Yes. I knew it [feliz].

[pausa]

Beo: Now it's your turn to be guesser.

Beo: I'm gonna pick an object, give me a second.

Beo olha para os objetos.

Beo: Okay, I'm ready. You can ask me five questions.

Usuário: Okay. Ah- is it rectangular?

Beo: No.

Usuário: Ahh- is it a cylinder?

Beo: No.

Usuário: Is it blue?

Beo: No.

Usuário: Ah- is it orange?

Beo: Yes.

Usuário: Is it squishy?

Beo: Yes.

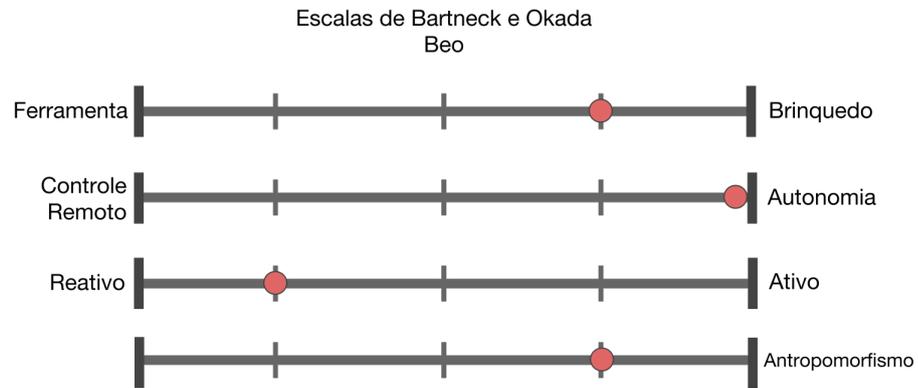
Beo: You're out of questions. Show me the object you're thinking of.

Usuário pega a Bola Lisa Laranja e mostra para o **Beo**.

Beo: Yes! That's the correct object [feliz].

Beo: Thanks for playing with me! See you next time.

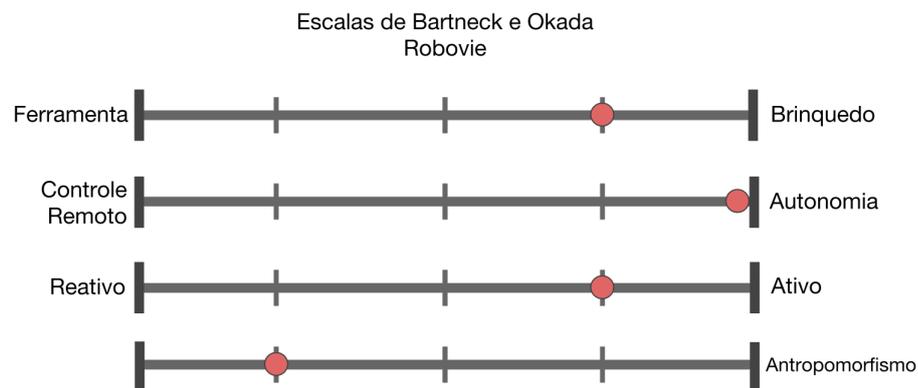
(a) Escalas de Bartneck e Okada (2001) para o robô **Beo**, utilizado neste trabalho.



(b) Escalas de Bartneck e Okada (2001) para o robô **NIMA** de Alemi, Meghdari e Ghazisaedy (2014).



(c) Escalas de Bartneck e Okada (2001) para o robô **Robovie** de Kanda et al. (2004).



Fonte: Autor.

4.1.2 Discussão das métricas de IHR e Escalas de Bartneck

A partir da transcrição do cenário (Seção 4.1.1) fica clara a implementação dos dois padrões de *design* de Kahn et al. (2008) anteriormente elencados. *Comunicação Didática* pode ser observado na Fase *Introdução* em que o robô explica o funcionamento do jogo para o usuário. Já o padrão *Tomando Turnos em um Contexto de Jogo* é mais explícito, servindo de fundação para as interações posteriores nas Fases *Beo Adivinhador* e *Usuário Adivinhador*.

A influência de cada um dos sete princípios de Goodrich e Jr (2003) é detalhada a seguir:

- Implicitamente alternar modos de autonomia e interfaces – **Beo** possui um único modo de autonomia (autonomia total), porém é necessário que se conecte a um *hardware* externo para executar o reconhecimento de objetos. Essa interface é realizada via **ROS**.
- Permitir que o robô faça uso de sinais humanos naturais – **Beo** é capaz de reconhecer fala e objetos e comunicar-se através de gestos, voz e expressões com os olhos.
- Manipular o mundo, ao invés do robô – O cenário de interação é realizado utilizando objetos físicos reais, que o usuário deve manipular para “mostrar” ao robô.
- Manipular a relação entre o robô e o mundo – O robô aponta para objetos físicos, além de demonstrar conhecer características físicas e visuais dos mesmos.
- Permitir que os usuário manipulem as informações apresentadas – Parte intrínseca do cenário de jogo. Os jogadores (tanto robô como usuário) devem lidar com informações adquiridas através das perguntas para realizar uma predição.
- Externalizar memória – As tarefas de reconhecimento de objetos e fala utilizam serviços externos ao robô, acessados através de uma interface de conexão.
- Ajudar os usuários a controlar a atenção – A própria proposta do cenário de interação é estimular e engajar os usuários a interagir com o robô para aprender inglês.

Finalmente, a estimativa da capacidades de IHR dos robôs segundo as escalas de Bartneck e Okada (2001) (Figuras 4.1a, 4.1b e 4.1c) surgiu do seguinte raciocínio:

- Ferramenta – Brinquedo: Os 3 robôs foram desenvolvidos com o intuito de estimular o aprendizado do inglês de forma divertida, por isso todos possuem algum elemento de *Brinquedo*. Entretanto, **NIMA** é utilizado para auxiliar um tutor humano durante aulas usuais, justificando sua tendência maior à *Ferramenta*, comparado com **Beo** e **Robovie**.

- Controle Remoto – Autonomia: **Beo** e **Robovie** operam de forma totalmente autônoma, sem intervenção humana, por isso sua colocação na extrema direita. Já **NIMA** recebe instruções de um humano (tutor) para iniciar roteiros de ações, resultando em sua posição central.
- Reativo – Ativo: **Beo** realiza suas rotinas de fala e movimentos em cada estado do cenário e aguarda uma resposta do usuário para prosseguir, resultando em sua tendência ao lado *Reativo* da escala. **NIMA** possui a mesma colocação pois opera de forma similar, tomando a iniciativa apenas em um de seus cenários propostos, quando os alunos fazem perguntas para o robô. Em contrapartida, **Robovie** tem uma forte tendência para o lado *Ativo* da escala, tomando a iniciativa na maior parte das interações, como por exemplo chamando os usuários humanos pelo nome.
- Antropomorfismo: Os 3 robôs pode ser classificados como robôs humanóides. Entretanto, existem algumas diferenças nas suas aparências e capacidades de expressão. **Robovie** conta com braços, porém não possui muitos detalhes faciais, sendo colocado à esquerda do centro. **NIMA** e **Beo** possuem aparências similares até certo ponto. Ambos tem uma aparência parecida com um humano pequeno, porém **Beo** possui olhos capazes de expressar algumas emoções simples, posicionando-o à direita de **NIMA**.

4.2 RESULTADOS PARA SISTEMA DE RECONHECIMENTO DE FALA

A Figura 4.2 e 4.3 apresentam a taxa de detecções falsas (falsos positivos) e a taxa de detecções corretas, respectivamente, para cada modelo de palavra-chave utilizado na Etapa 1 do sistema de reconhecimento de fala. O Quadro 4.1 apresenta a taxa de detecções falsas por minuto (FP/min) também para a Etapa 1, inferido a partir dos dados da Figura 4.3. O gráfico de barras da Figura 4.4 apresenta as detecções positivas da Etapa 3 do sistema de reconhecimento de fala para 15 vocalizações de cada característica possível dos objetos físicos. Finalmente, o Quadro 4.2 apresenta os tempos de resposta de cada etapa do sistema de reconhecimento de fala.

Quadro 4.1 – Taxa de detecções falsas por minuto para cada modelo de reconhecimento de voz em ambiente silencioso e barulhento.

Modelo	Ambiente	FP/min
Yes	Silencioso	0.00000
No	Silencioso	0.03333
Is It	Silencioso	0.00000
Yes	Barulhento	0.30000
No	Barulhento	0.05000
Is It	Barulhento	0.26666

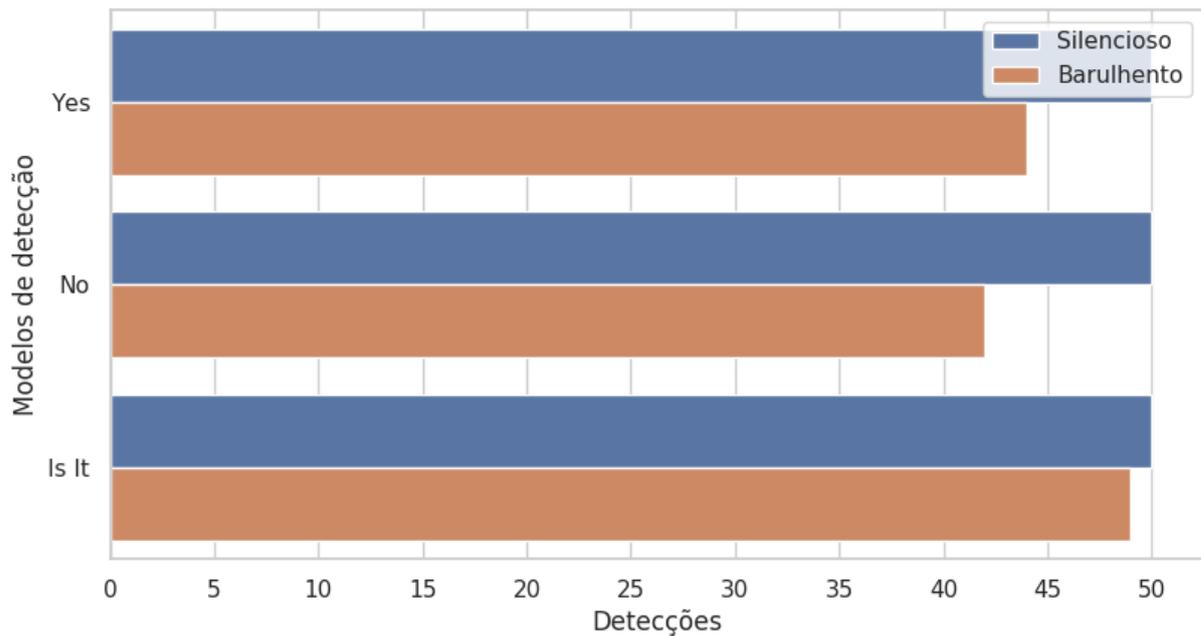
Fonte: Autor

Quadro 4.2 – Comparação dos tempos de resposta nas etapas do reconhecimento de fala.

Etapa	Tempo de Resposta Médio (ms)
Etapa 1 (Local)	258 ms
Etapa 2 + 3	2394 ms
Total	2652 ms

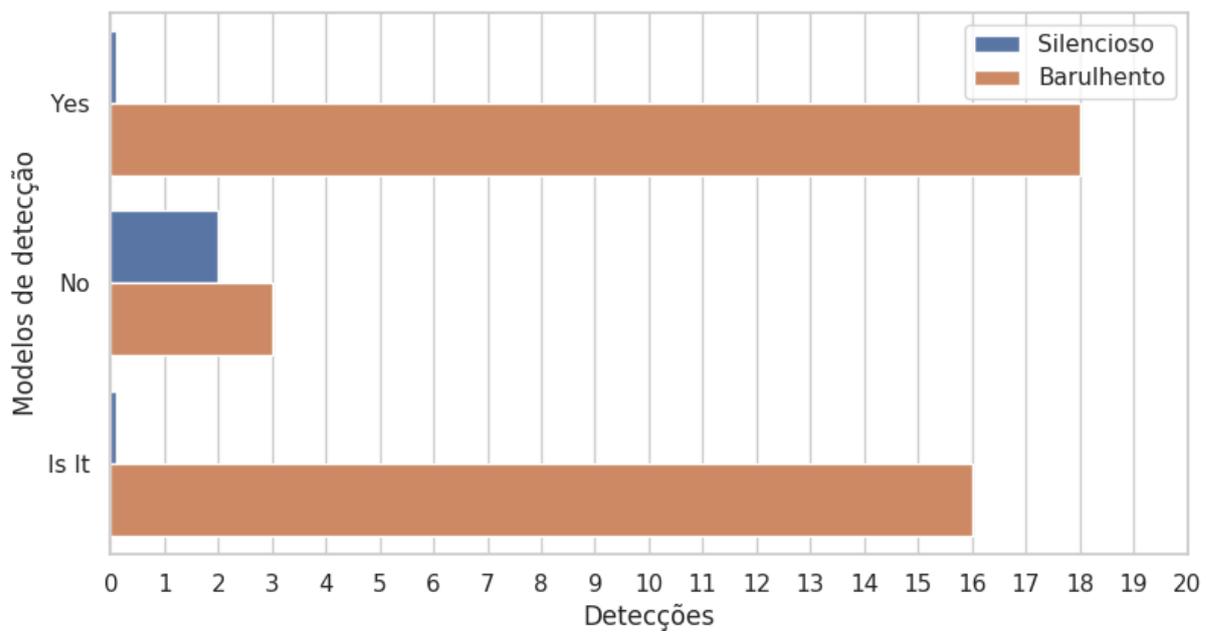
Fonte: Autor

Figura 4.2 – Detecções positivas para 50 verbalizações para cada modelo do sistema de reconhecimento de fala em ambiente silencioso e barulhento.



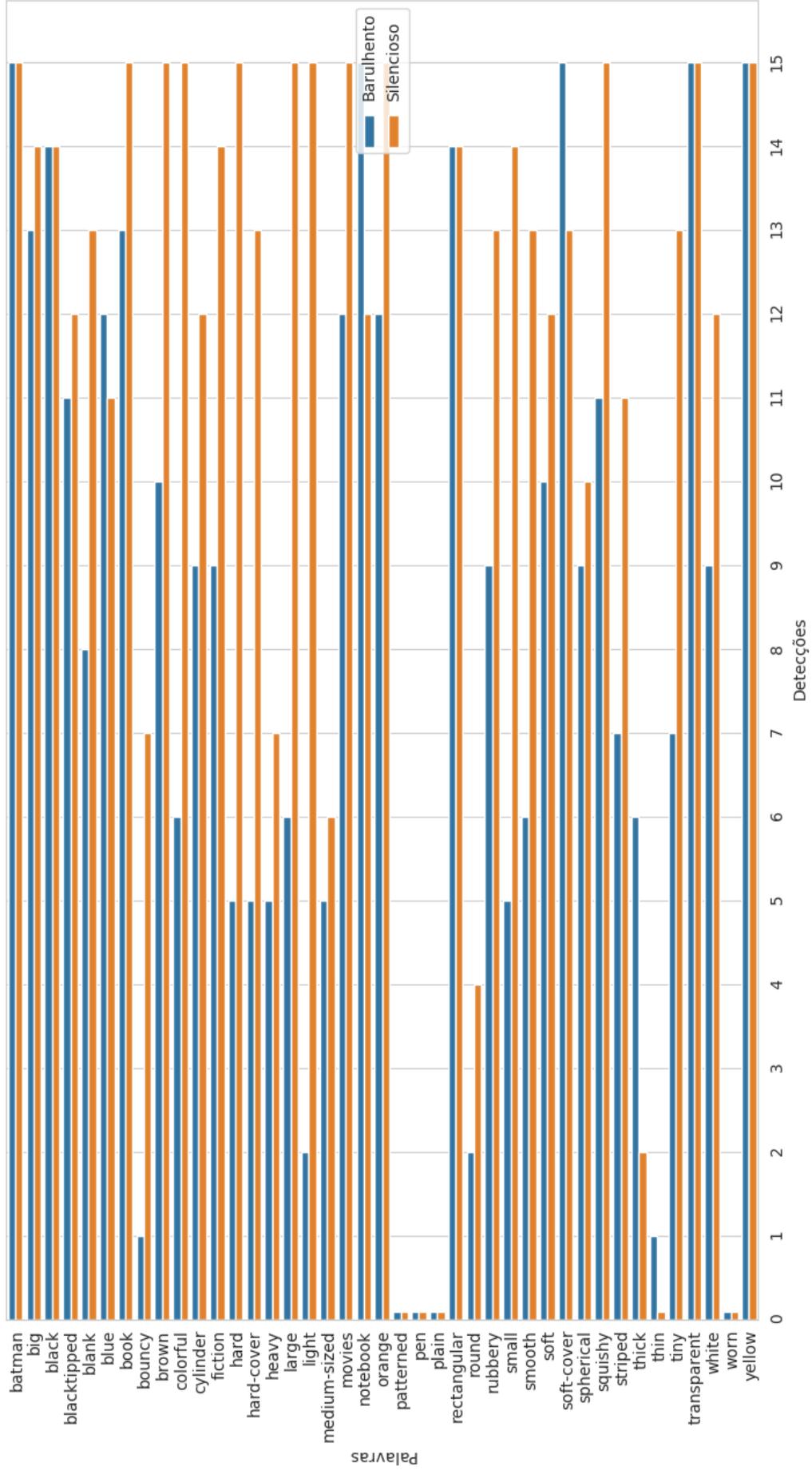
Fonte: Autor

Figura 4.3 – Detecções falsas em uma hora de gravação para cada modelo do sistema de reconhecimento (Etapa 1) de fala em ambientes silencioso e barulhento.



Fonte: Autor

Figura 4.4 – Detecções positivas para 15 verbalizações para cada palavra do sistema de reconhecimento de fala (Etapa 3) em ambientes silencioso e barulhento.



Fonte: Autor

4.2.1 Discussão dos resultados para o sistema de reconhecimento de fala

Observando o gráfico de barras da Figura 4.2 nota-se que a precisão de todos os modelos de reconhecimento de fala da Etapa 1 no ambiente silencioso foi de 100%. Já para o ambiente barulhento a precisão foi de 88% para o modelo “Yes”, 84% para “No” e 98% para “Is It”, resultando em uma precisão média de 90%. Analisando o gráfico da Figura 4.3 e o Quadro 4.1 nota-se que o único modelo com detecções falsas no ambiente silencioso foi “No”. Com um FP/min de 0.03333 pode-se inferir que uma detecção falsa provavelmente ocorrerá após 30 minutos. No ambiente barulhento, estima-se que uma detecção falsa ocorrerá após 3.333 minutos para o modelo “Yes”, 20 minutos para “No” e 3.75 minutos para “Is It”. Enquanto todos os modelos de detecção de palavra-chave da Etapa 1 conseguiram ótimos resultados, mesmo sob o ambiente barulhento, os resultados para a Etapa 3 foram mais discrepantes. Para a Etapa 3, observando gráfico de barras da Figura 4.4, podemos inferir que a precisão média para o ambiente Silencioso foi de 72,54% comparado a apenas 53,50% no ambiente Barulhento. Claramente, existe uma influência negativa de ruído na qualidade da detecção. Além disso, há uma grande discrepância na qualidade de detecção para palavras diferentes. Palavras monossilábicas (e.g. *pen*, *plain*, *thin*, *worn*) demonstraram grande dificuldade de detecção, com algumas sendo detectadas nem uma única vez. Analisando os tempos de resposta das etapas no Quadro 4.2 nota-se que a Etapa 1 possui um tempo muito melhor que a Etapa 2 e 3, mais de 2 segundos menor. Isso é coerente com o fato de que a detecção da Etapa 1 é realizada *offline* no próprio *SBC* do **Beo**, enquanto na Etapa 3 é realizada através de uma requisição via *internet*.

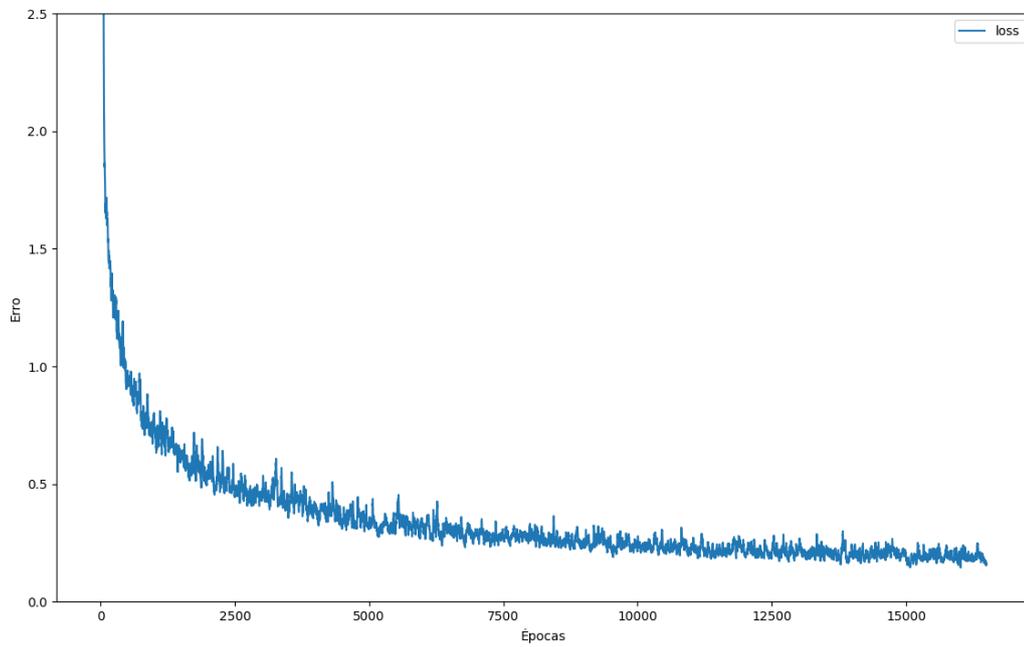
Levando em consideração que a maior parte dos estados do cenário de interação utiliza apenas a Etapa 1 do sistema de reconhecimento de fala e, tendo em vista os resultados observados nos gráficos pode-se afirmar que o sistema de reconhecimento de fala é robusto. Quando o ambiente é silencioso o sistema se mostra ótimo e, mesmo quando executado em um ambiente barulhento possui um desempenho acima do aceitável. Sua maior fraqueza é a Etapa 3, onde o tempo de resposta é relativamente alto, e a detecção de algumas palavras é difícil, principalmente em ambientes com ruídos.

4.3 RESULTADOS PARA O SISTEMA DE RECONHECIMENTO DE OBJETOS

A Figura 4.5 apresenta a variação do erro ao longo das épocas de treinamento para o conjunto de dados de treino (3458 imagens). Nas Figuras 4.6a, 4.6b e 4.6c é apresentada a *AP* para cada classe dos objetos físicos deste trabalho no conjunto de dados de validação (288 imagens). Na Figura 4.7 são apresentados o *mAP* e *IoU* médio também no conjunto de dados de validação. A Figura 4.8 apresenta os tempos de resposta para diferentes

esquemas de comunicação com o sistema de reconhecimento de objetos.

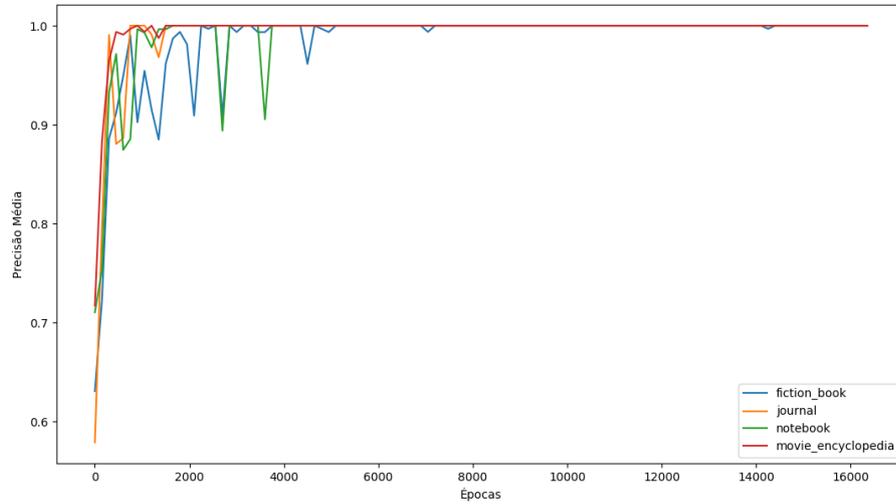
Figura 4.5 – Gráfico do erro ao longo do treinamento para o conjunto de dados de treino.



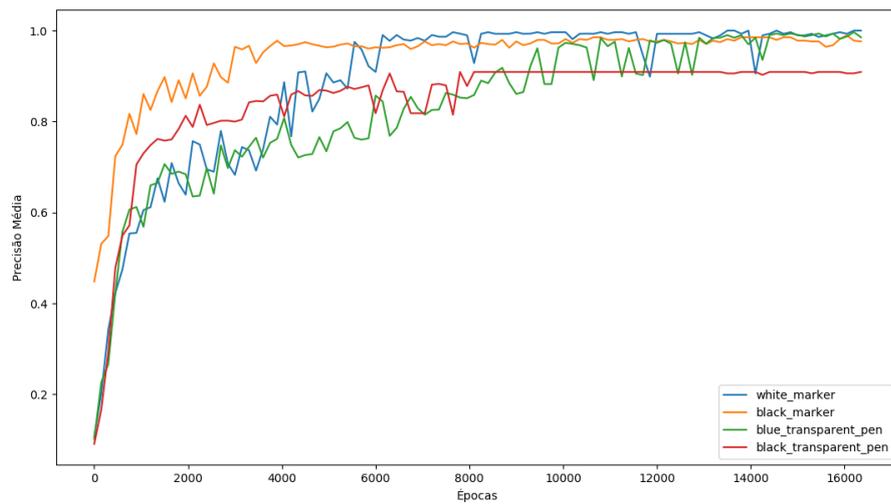
Fonte: Autor

Figura 4.6 – *AP* (Precisão Média) para cada classe de objetos no conjunto de dados de validação.

(a) *AP* (Precisão Média) por classe do tipo Caderno/Livro no conjunto de dados de validação.



(b) *AP* (Precisão Média) por classe do tipo Caneta no conjunto de dados de validação.



(c) *AP* (Precisão Média) por classe do tipo Bola no conjunto de dados de validação.

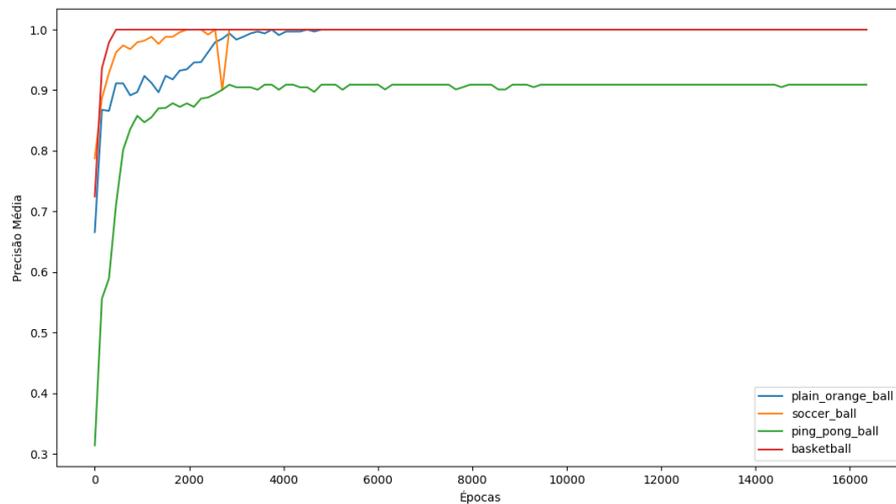
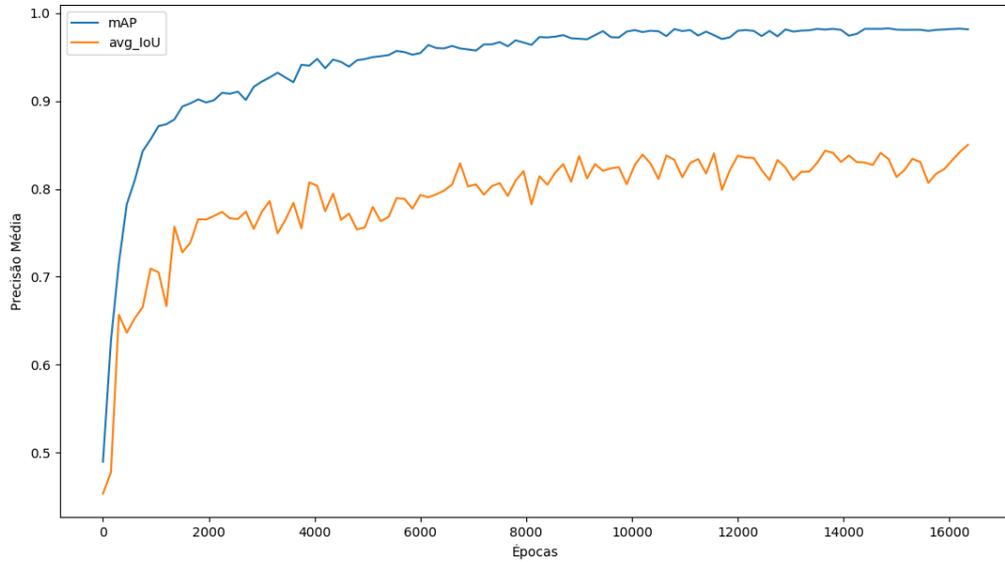
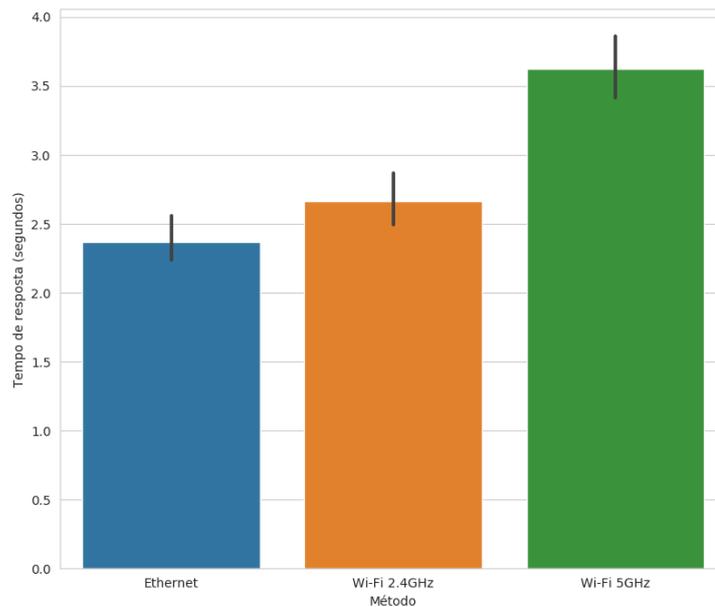


Figura 4.7 – *mAP* (Média das Precisões Médias) e *IoU* (Interseção sobre União) Médio no conjunto de dados de validação.



Fonte: Autor

Figura 4.8 – Comparação dos tempos de resposta para diferentes esquemas de comunicação com o sistema de reconhecimento de objetos.



Fonte: Autor

4.3.1 Discussão de resultados para o sistema de reconhecimento de objetos

Primeiramente, observando o gráfico da Figura 4.5 podemos observar que o treinamento ocorreu de forma usual para tais modelos, com o erro decresce rapidamente até atingir um platô. É interessante observar que apesar de não haver uma grande diminuição do valor do erro nas épocas posteriores do treinamento (Época > 7500) a continuação do mesmo ainda causa melhoras significativas no desempenho do modelo, como constatado especialmente pelo gráfico da Figura 4.6b, onde nota-se que o *AP* para a classe *blue_transparent_pen* (Caneta Azul Transparente) continua aumentando até as épocas finais do treinamento, bem após o platô do erro.

Pode-se constatar, observando o gráfico da Figura 4.7, que o modelo atingiu um ótimo desempenho. Com um *mAP* acima de 95% e *IoU* médio acima de 80% nos dados de validação. Ainda assim, nota-se que o desempenho varia entre as diferentes classes de objetos, em que as classes *black_transparent_pen* (Caneta Azul Transparente) e *ping_pong_ball* (Bola de Ping-Pong) foram as mais problemáticas. Uma possível explicação é que esses objetos são menores e possuem menos detalhes visuais (Figuras A.1 e A.6), dificultando sua detecção pelo modelo. Analisando o tempo de resposta, segundo o gráfico da Figura 4.8, nota-se que o método de conexão via cabo *Ethernet* obteve o menor tempo de resposta, com uma média de 2,37 segundos. Entretanto, quando comparando os métodos Wi-Fi, o tempo de resposta para o roteador Wi-Fi a 2.4GHz foi menor do que para o roteador de 5GHz: média de 2,66 segundos contra uma média de 3,63 segundos. Uma possível explicação para essa discrepância é a qualidade do roteador a 5GHz e possíveis interferências de outros eletrônicos transmitindo na mesma faixa de frequência.

Com alta precisão e um tempo de resposta razoável, pode-se dizer que o sistema de reconhecimento de objetos alcançou um desempenho bem acima do aceitável. Sua principal fraqueza no estado atual é a necessidade de um *hardware* dedicado, como a *NVIDIA Jetson TX2*, para sua execução.

5 CONSIDERAÇÕES FINAIS

A proficiência em língua inglesa tem sua importância amplificada no mundo globalizado em que vivemos hoje. Sua padronização informal como meio de comunicação internacional aumentou a busca por ferramentas eficazes no seu ensino. Robôs são cada vez mais capazes de interagir com humanos de maneiras complexas, e mostram um grande potencial para ensinar uma segunda língua.

Este trabalho apresentou o desenvolvimento de um cenário de interação com o robô **Beo** (Seção 3.1) para o ensino de língua inglesa. Inspirado nos conceitos de *gamification* e baseado em diretrizes de interação humano-robô, o trabalho consistiu na criação de um jogo de interação entre um humano e o robô com objetos físicos reais. O cenário desenvolvido é um jogo de adivinhação em que o robô e o usuário humano revezam-se nos papéis de adivinhador e escolhedor. O escolhedor escolhe algum dos objetos físicos e o adivinhador deve adivinhar o objeto escolhido após cinco perguntas.

Para que o robô seja capaz de executar tal cenário foi necessária a criação de um sistema de reconhecimento de fala e um sistema de reconhecimento de objetos. O sistema de reconhecimento de fala combina um modelo de detecção de palavras-chave *Snowboy*, com um tempo de resposta minúsculo, e um modelo de transcrição de fala na nuvem do *Google*, com um tempo de resposta maior. O sistema de reconhecimento de objetos consiste em um modelo de rede neural de convolução estado da arte (*Tiny YOLO V3*), executado em um *hardware* externo com interface realizada através de **ROS**.

Com o cenário e sistemas desenvolvidos, foram realizados diversos testes para analisar seu desempenho e tempo de resposta. Também foram realizadas análises posteriores quanto às capacidades de IHR de **Beo** e seu posicionamento segundo as escalas de Bartneck e Okada (2001) comparados a outros robôs de trabalhos relacionados. Os resultados apontam que ambos os sistemas de reconhecimento de fala e de objeto são robustos o suficiente para capacitar **Beo** a executar o cenário de interação proposto amplamente. Ainda assim, existem algumas limitações, como a necessidade de um *hardware* externo para o sistema de reconhecimento de objetos e *internet* para o sistema de reconhecimento de fala. Além disso, o treinamento para detecção de novos objetos é trabalhoso, e o sistema de reconhecimento de fala é falho em detectar certas palavras. Por limitações de tempo, também não foi possível realizar uma implementação mais profunda de elementos de *gamification*, a ser trabalhada futuramente.

Finalmente, levando em consideração os aspectos apresentados neste trabalho, é possível afirmar que robôs são capazes de interagir com o mundo e influenciar humanos de maneira cada vez mais complexa. Trabalhos futuros incluem principalmente realizar experimentos de interação com usuários e analisar seu impacto no aprendizado do inglês, motivação e percepção do robô, aplicando o questionário do Apêndice B ou similar.

Também podem ser exploradas técnicas para aumentar a complexidade do contexto de conversação do cenário, utilizando AIML, por exemplo. Adicionalmente, o desenvolvimento de mais cenários de interação, aplicando mais conceitos de *gamification* (e.g. recompensas por frequência de uso, conquistas, progressão, etc.) e análise de seus impactos no aprendizado é um caminho interessante a ser seguido.

REFERÊNCIAS BIBLIOGRÁFICAS

ALEMI, M.; MEGHDARI, A.; GHAZISAEDY, M. Employing humanoid robots for teaching english language in iranian junior high-schools. **International Journal of Humanoid Robotics**, World Scientific, v. 11, n. 03, p. 1450022, 2014.

ALEXANDER, C. **A pattern language: towns, buildings, construction**. [S.l.]: Oxford university press, 1977.

BARTNECK, C.; OKADA, M. Robotic user interfaces. In: CITESEER. **Proceedings of the Human and Computer Conference**. [S.l.], 2001. p. 130–140.

BLOOM, B. S. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. **Educational researcher**, Sage Publications Sage CA: Thousand Oaks, CA, v. 13, n. 6, p. 4–16, 1984.

BRIDGELAND, J. M.; JR, J. J. D.; MORISON, K. B. The silent epidemic: Perspectives of high school dropouts. **Civic Enterprises**, ERIC, 2006.

CLASSDOJO. **ClassDojo**. 2018. Disponível em: <<https://www.classdojo.com/>>.

DELQUADRI, J. et al. Classwide peer tutoring. **Exceptional children**, SAGE Publications Sage CA: Los Angeles, CA, v. 52, n. 6, p. 535–542, 1986.

DETERDING, S. Gamification: designing for motivation. **interactions**, ACM, v. 19, n. 4, p. 14–17, 2012.

DETERDING, S. et al. From game design elements to gamefulness: defining gamification. In: ACM. **Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments**. [S.l.], 2011. p. 9–15.

DICHEVA, D. et al. Gamification in education: A systematic mapping study. **Journal of Educational Technology & Society**, JSTOR, v. 18, n. 3, 2015.

DUOLINGO. **Duolingo**. 2018. Disponível em: <<https://en.duolingo.com/>>.

EDMODO. **Edmodo**. 2018. Disponível em: <<https://www.edmodo.com/>>.

FLORES, J. F. F. Using gamification to enhance second language learning. **Digital Education Review**, Digital Education Observatory (OED), n. 27, p. 32–54, 2015.

GLOVER, I. Play as you learn: Gamification as a technique for motivating learners. In: HERRINGTON, J.; COUROS, A.; IRVINE, V. (Ed.). **Proceedings of EdMedia + Innovate Learning 2013**. Victoria, Canada: Association for the Advancement of Computing in Education (AACE), 2013. p. 1999–2008. Disponível em: <<https://www.learntechlib.org/p/112246>>.

GOODRICH, M.; JR, D. R. O. Seven principles of efficient human robot interaction. In: **IEEE International Conference on Systems Man and Cybernetics**. [S.l.: s.n.], 2003. v. 4, p. 3943–3948.

GOOGLE. **Google Cloud Speech-to-Text**. 2018. Disponível em: <<https://cloud.google.com/speech-to-text/>>.

GORHAM, J. The relationship between verbal teacher immediacy behaviors and student learning. **Communication education**, Taylor & Francis Group, v. 37, n. 1, p. 40–53, 1988.

GROH, F. Gamification: State of the art definition and utilization. **Institute of Media Informatics Ulm University**, v. 39, p. 31, 2012.

HAMARI, J.; KOIVISTO, J.; SARSA, H. Does gamification work?—a literature review of empirical studies on gamification. In: IEEE. **2014 47th Hawaii international conference on system sciences (HICSS)**. [S.l.], 2014. p. 3025–3034.

HELLOTALK. **HelloTalk**. 2018. Disponível em: <<https://www.hellotalk.com/>>.

HENKEMANS, O. A. B. et al. Using a robot to personalise health education for children with diabetes type 1: A pilot study. **Patient education and counseling**, Elsevier, v. 92, n. 2, p. 174–181, 2013.

HUBBARD, P.; LEVY, M. The scope of call education. **Teacher education in CALL**, p. 3–20, 2006.

HUOTARI, K.; HAMARI, J. Defining gamification: a service marketing perspective. In: ACM. **Proceeding of the 16th international academic MindTrek conference**. [S.l.], 2012. p. 17–22.

KAHN, P. H. et al. Design patterns for sociality in human-robot interaction. In: ACM. **Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction**. [S.l.], 2008. p. 97–104.

KANDA, T. et al. Interactive robots as social partners and peer tutors for children: A field trial. **Human–Computer Interaction**, Taylor & Francis, v. 19, n. 1-2, p. 61–84, 2004.

KENNEDY, J. et al. Social robot tutoring for child second language learning. In: IEEE PRESS. **The Eleventh ACM/IEEE International Conference on Human Robot Interaction**. [S.l.], 2016. p. 231–238.

KIRYAKOVA, G.; ANGELOVA, N.; YORDANOVA, L. Gamification in education. In: PROCEEDINGS OF 9TH INTERNATIONAL BALKAN EDUCATION AND SCIENCE CONFERENCE. [S.l.], 2014.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2012. p. 1097–1105.

KRUIJFF-KORBAYOVA, I. et al. Effects of off-activity talk in human-robot interaction with diabetic children. In: IEEE. **Robot and Human Interactive Communication, 2014 ROMAN: The 23rd IEEE International Symposium on**. [S.l.], 2014. p. 649–654.

KUHL, P. K. Early language acquisition: cracking the speech code. **Nature reviews neuroscience**, Nature Publishing Group, v. 5, n. 11, p. 831, 2004.

_____. Is speech learning gated by the social brain? **Developmental science**, Wiley Online Library, v. 10, n. 1, p. 110–120, 2007.

LEE, J. J.; HAMMER, J. Gamification in education: What, how, why bother? **Academic exchange quarterly**, Chattanooga State Technical Community College, v. 15, n. 2, p. 146, 2011.

LEYZBERG, D. et al. The physical presence of a robot tutor increases cognitive learning gains. In: **Proceedings of the Annual Meeting of the Cognitive Science Society**. [S.l.: s.n.], 2012. v. 34, n. 34.

MINDSNACKS. **MindSnacks**. 2018. Disponível em: <<http://mindsnacks.com/>>.

OSIPOV, I. V. et al. Study of gamification effectiveness in online e-learning systems. **International Journal of advanced computer science and applications**, v. 6, n. 2, p. 71–77, 2015.

REDMON, J.; FARHADI, A. YOLO9000: better, faster, stronger. **CoRR**, abs/1612.08242, 2016. Disponível em: <<http://arxiv.org/abs/1612.08242>>.

SCHUMANN, J. H. Where is cognition?: Emotion and cognition in second language acquisition. **Studies in second language acquisition**, Cambridge University Press, v. 16, n. 2, p. 231–242, 1994.

SIMÕES, J.; REDONDO, R. D.; VILAS, A. F. A social gamification framework for a k-6 learning platform. **Computers in Human Behavior**, Elsevier, v. 29, n. 2, p. 345–353, 2013.

SNOWBOY. **Snowboy Hotword Detector**. 2018. Disponível em: <<https://snowboy.kitt.ai/>>.

STONE, R. **Rosetta Stone**. 2018. Disponível em: <<https://www.rosettastone.com/>>.

VESSELINOV, R.; GREGO, J. Duolingo effectiveness study. **City University of New York, USA**, v. 28, 2012.

WERBACH, K. (re) defining gamification: A process approach. In: SPRINGER. **International conference on persuasive technology**. [S.l.], 2014. p. 266–272.

WITT, P. L.; WHEELLESS, L. R.; ALLEN, M. A meta-analytical review of the relationship between teacher immediacy and student learning. **Communication Monographs**, Taylor and Francis Ltd, v. 71, n. 2, p. 184–207, 2004.

YBARRA, R.; GREEN, T. Using technology to help esl/efl students develop language skills. **The Internet TESL Journal**, v. 9, n. 3, p. 1–5, 2003.

APÊNDICE A – LISTA COMPLETA DOS OBJETOS FÍSICOS

Este apêndice apresenta a lista completa dos objetos físicos utilizados no jogo de interação. Para cada objeto, é apresentada com uma foto do objeto físico real e uma lista de todas as possíveis características que podem ser mencionadas durante o jogo de interação.

No total foram escolhidos 12 objetos de 3 categorias distintas: bolas, canetas e livros/cadernos. Os objetos são apresentados a seguir com seus nome em inglês – como o objeto é mencionado dentro do jogo – seguido de sua tradução para português.

A.1 – FOTOGRAFIAS DOS OBJETOS



Figura A.1 – **Blue Transparent Pen** - Caneta Azul Transparente



Figura A.2 – **Black Transparent Pen** - Caneta Preta Transparente



Figura A.3 – **Black Marker** - Marcador Preto



Figura A.4 – **White Marker** - Marcador Branco



Figura A.5 – **Soccer Ball** - Bola de Futebol



Figura A.6 – **Ping Pong Ball** - Bolinha de Ping Pong



Figura A.7 – **Basketball** - Bola de Basquete



Figura A.8 – **Plain Orange Ball** - Bola Lisa Laranja

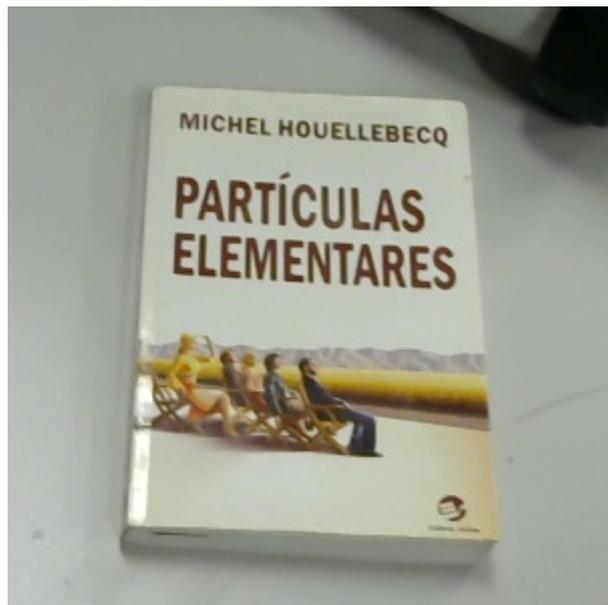


Figura A.9 – **Fiction Book** - Livro de Ficção

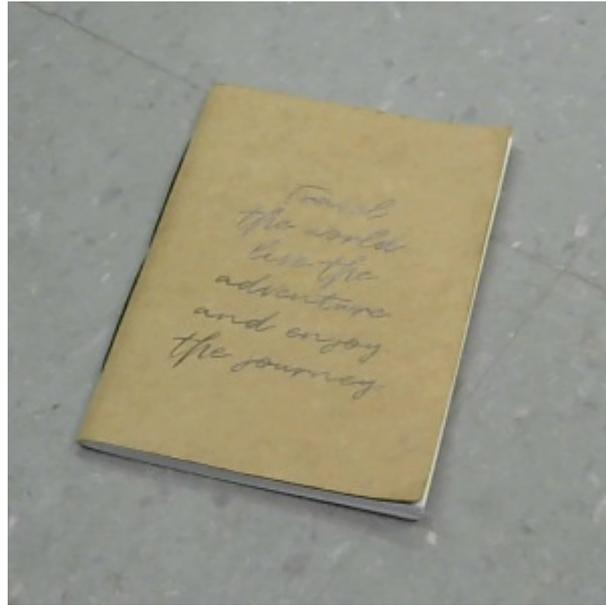


Figura A.10 – **Journal** - Diário

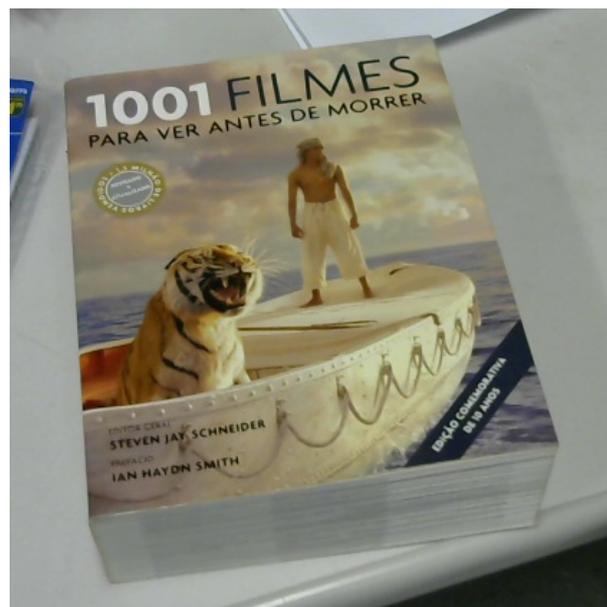


Figura A.11 – **Movie Encyclopedia** - Enciclopédia de Filmes

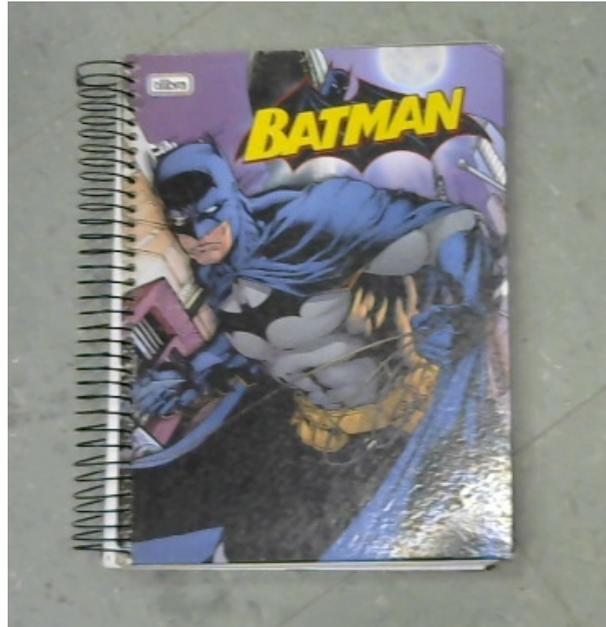


Figura A.12 – **Batman Notebook** - Caderno do Batman

A.2 – LISTA COMPLETA DAS CARACTERÍSTICAS DOS OBJETOS

- **Blue Transparent Pen (Caneta Azul Transparente – Figura A.1):** transparent, blue, cylinder, thin, pen, smooth
- **Black Transparent Pen (Caneta Preta Transparente – Figura A.2):** transparent, black, light, cylinder, thin, pen, rubbery
- **Black Marker (Marcador Preto – Figura A.3):** black, blue, cylinder, light, thick, smooth
- **White Marker (Marcador Branco – Figura A.4):** white, black tipped, light, cylinder, thick, smooth
- **Soccer Ball (Bola de Futebol – Figura A.5):** hard, striped, round, spherical, blue, white, patterned, medium-sized, light
- **Ping Pong Ball (Bola de Ping Pong – Figura A.6):** tiny, hard, bouncy, round, spherical, plain, light

- **Basketball (Bola de Basquete – Figura A.7):** large, big, round, spherical, bouncy, blue, striped, heavy, worn
- **Plain Orange Ball (Bola Lisa Laranja – Figura A.8):** plain, orange, squishy, soft, small, round, spherical, worn
- **Fiction Book (Livro de Ficção – Figura A.9):** white, book, yellow, medium-sized, fiction, rectangular, smooth, light
- **Journal (Diário – Figura A.10):** brown, thin, blank, soft-cover, notebook, light, small, rectangular
- **Movie Encyclopedia (Enciclopédia de Filmes – Figura A.11):** colorful, book, thick, heavy, movies, large, big
- **Batman Notebook (Caderno do Batman – Figura A.12):** colorful, batman, hard-cover, medium-sized, rectangular, notebook, light, large, big

APÊNDICE B – QUESTIONÁRIO DE ANÁLISE QUALITATIVA DO CENÁRIO DE INTERAÇÃO

Incluído na próxima página, devido à formatação.

Questionário de Análise Qualitativa do Cenário de Interação

Esse questionário faz parte do trabalho de conclusão de curso “Um Jogo de Interação Humano-Robô para o Ensino de Língua Inglesa”. Responda as perguntas a seguir referentes aos aspectos do cenário de interação realizado com o robô **Beo**.

Sobre você.

Responda os itens abaixo com informações sobre você mesmo.

1. Nome: _____
2. Idade: _____
3. Grau de Formação Acadêmica: <input type="radio"/> Ensino Fundamental Completo <input type="radio"/> Ensino Médio Completo <input type="radio"/> Ensino Superior Completo <input type="radio"/> Pós-Graduação Completa
4. Você já interagiu com algum robô ou avatar virtual anteriormente? <input type="radio"/> Sim. <input type="radio"/> Não.
5. Você já interagiu com algum robô fisicamente presente anteriormente? <input type="radio"/> Sim. <input type="radio"/> Não.
6. Qual seu grau de familiaridade com robótica? Nenhum <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> Elevado

Sobre o Robô.

Responda as perguntas abaixo sobre o robô e a experiência de interação.

7. Você se sentiu confortável durante a interação com o robô? Pouco <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> Muito
8. Você achou o robô inteligente? Pouco <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> Muito
9. Você achou o robô irritante? Pouco <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> Muito
10. Você se sentiu estimulado a interagir com o robô? Pouco <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> Muito
11. O quão divertida você achou a interação com o robô? Pouco <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> Muito
12. Você teve dificuldade em compreender o que o robô falava? Pouco <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> Muito
13. Você acha que o robô teve dificuldades em compreender o que o você falava? Pouco <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> Muito
14. Você acha que a interação com o robô o estimulou no aprendizado de inglês? Pouco <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> Muito
15. O quanto você acha que o robô tomou a iniciativa na interação? Pouco <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> — <input type="radio"/> Muito

16. Dada a oportunidade, com que frequência você gostaria de interagir com o robô?

- Nunca mais
- Uma vez por semana
- Duas a três vezes por semana
- Mais de três vezes por semana
- Diariamente

APÊNDICE C – TRANSCRIÇÃO DE INTERAÇÃO EM PORTUGUÊS

Para completude deste trabalho, neste apêndice é apresentado a transcrição da interação da Seção 4.1.1 traduzida para português.

Beo: Olá [acenando]! Quer jogar um jogo [piscadela]?

Beo: Esse será um jogo de adivinhação com os objetos em nossa frente [apontando para os objetos].

Beo: Cada um de nós irá escolher um objeto, enquanto a outra pessoa tem que adivinhar que objeto é.

Beo: Nós vamos fazer isso fazendo perguntas de “Sim” e “Não” sobre as características desses objetos.

Beo: Todas as nossas perguntas devem começar com “Ele é”. Por exemplo, “Ele é grande?”, “Ele é branco?” ou “Ele é um cilindro?”.

Beo: Conseguiu entender tudo até aqui?

Usuário: Sim.

Beo: Ótimo! Vamos começar.

Beo: Ok. Eu vou começar como o adivinhador. Escolha um dos objetos na minha frente. Mas não me conte.

[pausa]

Usuário sinaliza que escolheu o objeto Bola de Basquete.

Beo: Quando estiver pronto, diga “Sim”.

Usuário: Sim.

Beo: Ok. Eu vou tentar adivinhar o objeto que você escolheu.

Beo: Responda minhas perguntas!

Beo: Ele é retangular?

Usuário: Não.

Beo: Ele é uma caneta?

Usuário: Não.

Beo: Ele é branco?

Usuário: Não.

Beo: Ele é redondo?

Usuário: Sim.

Beo: Ele quica?

Usuário: Sim.

Beo: Você está pensando na bola de basquete, certo?

Usuário: Sim.

Beo: Sim. Eu sabia [feliz].

[pausa]

Beo: Agora é sua vez de ser o adivinhador.

Beo: Eu vou escolher o objeto, me dê um segundo.

Beo olha para os objetos.

Beo: Ok, estou pronto. Pode fazer cinco perguntas.

Usuário: Ok. Ah- ele é retangular?

Beo: Não.

Usuário: Ahh- ele é um cilindro?

Beo: Não.

Usuário: Ele é azul?

Beo: Não.

Usuário: Ah- ele é laranja?

Beo: Sim.

Usuário: Ele é apertável?

Beo: Sim.

Beo: Acabaram suas perguntas. Me mostre o objeto que você está pensando.

Usuário pega a Bola Lisa Laranja e mostra para o **Beo**.

Beo: Sim! Esse é o objeto certo [feliz].

Beo: Obrigado por jogar comigo! Te vejo na próxima vez.