

UNIVERSIDADE FEDERAL DE SANTA MARIA  
CENTRO DE TECNOLOGIA  
CURSO DE GRADUAÇÃO EM ENGENHARIA DE CONTROLE E  
AUTOMAÇÃO

Roberta de Mendonça Pretto

**GERAÇÃO DE APRESENTAÇÕES ORAIS COM EXPRESSIVIDADE  
EMOCIONAL EM UM ROBÔ HUMANOIDE**

Santa Maria, RS  
2022

**Roberta de Mendonça Preto**

**GERAÇÃO DE APRESENTAÇÕES ORAIS COM EXPRESSIVIDADE EMOCIONAL EM UM ROBÔ HUMANOIDE**

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

ORIENTADOR: Prof. Rodrigo da Silva Guerra

Santa Maria, RS  
2022

---

©2022

Todos os direitos autorais reservados a Roberta de Mendonça Pretto. A reprodução de partes ou do todo deste trabalho só poderá ser feita mediante a citação da fonte.

Endereço: Rua Alcides Ribeiro, 89, Tupanciretã, RS

Fone (0xx) +55 55 984713870; End. Eletr.: [betampretto93@gmail.com](mailto:betampretto93@gmail.com)

**Roberta de Mendonça Pretto**

**GERAÇÃO DE APRESENTAÇÕES ORAIS COM EXPRESSIVIDADE EMOCIONAL EM  
UM ROBÔ HUMANOIDE**

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

**Aprovado em 28 de janeiro de 2022:**

---

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

---

**George França dos Santos, Dr. (UFT)**

---

**Daniel Fernando Tello Gamarra, Dr. (UFSM)**

Santa Maria, RS  
2022

## DEDICATÓRIA

*A minha mãe Inára e a todos os profissionais da educação pública de qualidade*

## **AGRADECIMENTOS**

*Acima de tudo, gostaria de agradecer à minha família pelo apoio incondicional que me foi dado durante todo o percurso da graduação. Aos meus colegas e amigos da Qiron Robotics agradeço pelos ensinamentos e pelo crescimento que me proporcionaram.*

*Aos Professores, um agradecimento especial ao Professor Rodrigo da Silva Guerra pelos anos de orientação e ensinamentos, à Professora Soraia Napoleão Freitas, por me mostrar como as diferenças podem ser enriquecedoras e ao Professor Rafael Concatto Beltrame, um exemplo de educador dentro da UFSM.*

*Aos amigos, agradeço à Audren, que tanto me ajudou e incentivou durante as disciplinas e trabalhos em grupo, o caminho foi longo, mas nós conseguimos. Agradeço ao Moisés que, além da amizade, me forneceu parte do material necessário para a confecção deste trabalho.*

*O meu mais sincero obrigada a todos aqueles que acreditaram em mim, me apoiaram, me ensinaram e me fizeram crescer de alguma forma.*

*O infinito é realmente  
Um dos deuses mais lindos*

*(Legião Urbana)*

## RESUMO

### GERAÇÃO DE APRESENTAÇÕES ORAIS COM EXPRESSIVIDADE EMOCIONAL EM UM ROBÔ HUMANOIDE

AUTORA: Roberta de Mendonça Pretto  
ORIENTADOR: Rodrigo da Silva Guerra

Com o aumento das atividades realizadas de forma remota, sistemas de auto-atendimento e, mais recentemente, a necessidade de distanciamento social devido à pandemia de COVID-19, o uso da robótica e sistemas autônomos para atividades que demandam interações com o público têm sido cada vez mais estudado. Entre as soluções emergentes para essa demanda estão os robôs sociais e, em particular, os robôs humanoides. Esses últimos, através de sua forma antropomórfica, buscam comunicar-se com humanos de forma mais intuitiva e natural, causando uma sensação de presença social, alcançando assim um melhor engajamento com as pessoas. Dentro desse contexto, um problema extremamente relevante é a expressão de gestos e movimentos naturais, que estejam em sintonia com a fala do robô. O ato de demonstrar emoções em fala aumenta a comunicação afetiva e a compreensão do que é dito através da empatia entre o robô e público alvo. Neste trabalho propomos um método para a geração automática de apresentações orais, a partir de roteiros escritos, com movimentos corporais e expressões, baseadas nas emoções automaticamente associadas ao texto, via algoritmo treinado com redes neurais artificiais.

**Palavras-chave:** Robô Humanoide. Geração de Movimentos. Comunicação Não Verbal. Linguagem Corporal. Emoções



## ABSTRACT

### GENERATION OF PRESENTATIONS WITH EMOTIONAL EXPRESSIVENESS IN A HUMANOID ROBOT

AUTHOR: Roberta de Mendonça Pretto

ADVISOR: Rodrigo da Silva Guerra

With the increase in activities performed remotely, self-service systems and, more recently, the need for social distancing due to the COVID-19 pandemic, the use of robotics and autonomous systems for activities that demand interactions with the public have been increasingly studied. Among the emerging solutions to this demand are social robots and, in particular, humanoid robots. The latter, through their anthropomorphic form, seek to communicate with humans in a more intuitive and natural way, causing a sense of social presence, thus achieving a better engagement with people. Within this context, an extremely relevant problem is the expression of natural gestures and movements, which are in tune with the robot's speech. The act of demonstrating emotions in speech increases affective communication and understanding of what is said by generating empathy between the robot and its target audience. In this work we propose a method for the automatic generation of oral presentations, from written scripts, with body movements and expressions, based on emotions automatically associated with the text, via algorithm trained with artificial neural networks.

**Keywords:** Humanoid Robot. Generation of Movements. Non verbal communication. Body language. Emotions

## LISTA DE FIGURAS

Figura 1.1 – Kaspar sendo controlado por um pesquisador. ....	13
Figura 1.2 – Robô Keepon e sua constituição. ....	14
Figura 1.3 – PARO, o robô terapêutico. ....	15
Figura 1.4 – AIBO, o cão robô. ....	15
Figura 1.5 – Pepper. ....	16
Figura 1.6 – Visão geral do Kismet e seus componentes. ....	17
Figura 2.1 – Face do robô DoRIS ....	21
Figura 2.2 – Raiva (1), nojo (2), medo (3), surpresa (4), felicidade (5), tristeza (6) e desprezo (7) ....	22
Figura 2.3 – Robô WE-4RII expressando diferentes emoções ....	24
Figura 2.4 – Robô NAO ....	25
Figura 2.5 – Figuras apresentadas por Fritz Heider e Maianne Simmel ....	26
Figura 2.6 – Movimentos gerados por entrada de voz ....	27
Figura 2.7 – Emoções representadas em um plano 2-D ....	29
Figura 3.1 – Emoções de Ekman em um plano 2-D ....	31
Figura 3.2 – O robô Beo ....	33
Figura 3.3 – Graus de liberdade da movimentação da cabeça do Beo ....	34
Figura 3.4 – Servo-atuador XL-320 ....	35
Figura 3.5 – <i>ASUS Tinker Board</i> ....	36
Figura 3.6 – <i>Olhos para as expressões de alegria, neutralidade, surpresa, raiva, tristeza e medo</i> ....	37
Figura 3.7 – Ambiente para criação de Movimentos ....	38
Figura 3.8 – Alegria: quadro a quadro ....	39
Figura 3.9 – Surpresa: quadro a quadro ....	39
Figura 3.10 – Raiva: quadro a quadro ....	40
Figura 3.11 – Medo: quadro a quadro ....	40
Figura 3.12 – Tristeza: quadro a quadro ....	40
Figura 3.13 – Neutro ....	41
Figura 3.14 – Fluxograma do algoritmo para detecção automática da emoção. ....	42
Figura 3.15 – Fluxograma do algoritmo com inserção manual da valência e excitação. ....	43
Figura 3.16 – Posição zero do robô ....	44
Figura 4.1 – Apresentação do Questionário ....	45
Figura 4.2 – Apresentação do Questionário ....	46
Figura 5.1 – Matriz de Confusão ....	48
Figura 5.2 – Taxa de reconhecimento das emoções ....	48
Figura 5.3 – Taxa de reconhecimento da emoção surpresa ....	49
Figura 5.4 – Taxa de reconhecimento da emoção medo ....	49
Figura 5.5 – Classificação de frases transpostas para um plano 2-D ....	50
Figura 5.6 – Classificação de frases transpostas para um plano 2-D ....	51
Figura 5.7 – Classificação de frases por região do gráfico ....	51
Figura 5.8 – Aplicação gráfica ....	52
Figura 5.9 – Composição da Apresentação ....	53

## LISTA DE QUADROS

Quadro 2.1 – Características faciais de cada emoção .....	23
Quadro 3.1 – Quadro com os ângulos de referência de cada emoção .....	32
Quadro 3.2 – Características do Servo Motor XL-320 .....	35

## LISTA DE ABREVIATURAS E SIGLAS

<i>HTML</i>	Hyper Text Markup Language
<i>LIBRAS</i>	Língua Brasileira de Sinais
<i>ME</i>	Ministério da Economia
<i>SGD</i>	Secretaria de Governo Digital
<i>UFPB</i>	Universidade Federal da Paraíba
<i>UFSM</i>	Universidade Federal de Santa Maria

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>12</b>
1.1	CONTEXTUALIZAÇÃO .....	12
1.2	JUSTIFICATIVA.....	17
1.3	ESTRUTURA DO TRABALHO .....	19
1.4	OBJETIVOS .....	19
1.4.1	<b>Objetivos Gerais</b> .....	<b>19</b>
1.4.2	<b>Objetivos Específicos</b> .....	<b>20</b>
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA</b> .....	<b>21</b>
2.1	TRABALHOS RELACIONADOS:.....	21
<b>3</b>	<b>MATERIAIS E MÉTODOS</b> .....	<b>30</b>
3.1	DETECÇÃO DA EMOÇÃO EM TEXTO .....	30
3.2	DESCRIÇÃO DE HARDWARE .....	32
3.3	COMPOSIÇÃO ESTÉTICA PARA REPRESENTAÇÃO DE EMOÇÕES .....	36
3.4	DESCRIÇÃO DO ALGORITMO .....	41
<b>4</b>	<b>QUESTIONÁRIO</b> .....	<b>45</b>
<b>5</b>	<b>RESULTADOS</b> .....	<b>47</b>
5.1	RESULTADO ESTÉTICO .....	47
5.2	MÉTODO PARA DETECÇÃO DE EMOÇÃO EM TEXTO .....	50
5.3	LOCAL DE APLICAÇÃO .....	52
<b>6</b>	<b>CONCLUSÃO</b> .....	<b>54</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>56</b>

# 1 INTRODUÇÃO

Com o crescente aumento das atividades realizadas de forma remota, o emprego de sistemas de auto-atendimento e o uso da robótica e sistemas autônomos para atividades que demandam interações com o público têm sido cada vez mais estudado. Não é incomum, ao andar por shoppings, ruas, ou estabelecimentos comerciais, encontrar postos de autoatendimento ou publicidade, cuja premissa principal consiste na criação de um personagem que cativa o público e chame a sua atenção para produtos ou serviços específicos de alguma empresa. Esse personagem geralmente precisa ser capaz de se comunicar, ser simpático, gerar empatia e atrair as pessoas que estão ao seu redor. Dessa forma, neste capítulo iremos apresentar a contextualização e as motivações que norteiam este trabalho.

## 1.1 CONTEXTUALIZAÇÃO

Nos últimos anos houve um grande aumento da presença dos assistentes virtuais, como a Alexa da Amazon, a Siri da Apple e a assistente pessoal da Google. Esses assistentes são capazes de gerenciar casas inteligentes, agendar reuniões e até mesmo realizar ligações automáticas, tudo de forma natural e interativa. Esses assistentes porém, não apresentam formas antropomórficas. A aparência humana é essencial para a ativação do chamado sistema espelho (RIZZOLATTI; ARBIB, 1998), que é um sistema composto por neurônios responsáveis por reconhecer e reproduzir, mentalmente, certos comportamentos observados em terceiros, como se estivessem sendo realizados pelo próprio observador. Esse sistema forma um alicerce fundamental relacionando a empatia e a capacidade de observar, imitar e aprender observando outro, e é essencial para o desenvolvimento e aprendizagem da linguagem, comunicação e comportamento humano (RIZZOLATTI; CRAIGHERO, 2004).

No Brasil, um dos mais conhecidos exemplos de personagem digital de aparência humanoide que é capaz de gerar gestos é trazido no aplicativo de acessibilidade em língua dos sinais, *Hand Talk*. O *Hand Talk* é um tradutor para Libras, produzido pela empresa *Hand Talk*, e tem como intérprete o HUGO, que é uma espécie de cartoon 3D animado que representa a figura de um homem capaz de realizar movimento, com mãos e braços, de forma a traduzir textos e áudios para a língua brasileira de sinais (LIBRAS) através de inteligência artificial. Na mesma área também podemos citar o VLibras, resultado da parceria entre o Ministério da Economia (ME), por meio da Secretaria de Governo Digital (SGD) e a Universidade Federal da Paraíba (UFPB). Mas em ambos estes casos, estes são apenas personagens virtuais representados graficamente em um tela digital.

As interações sociais são extremamente importantes para o desenvolvimento de

robôs sociais. A comunicação não verbal, através de gestos e movimentos, é essencial para a contextualização do objetivo do robô, seja ele informar, cuidar ou ensinar. Movimentos expressivos e adequados fazem com que o robô deixe de ser percebido como uma "máquina" e se torne um "ser" com o qual uma pessoa pode criar uma relação de empatia, harmonizando as interações entre eles. Essa relação melhora o desempenho do robô na sua função trazendo benefícios para as pessoas em diversas áreas, como educação, cuidados hospitalares e, até mesmo, ajudando pessoas com dificuldades em interações sociais, como o autismo.

O uso de robôs tem sido amplamente difundido também nas áreas educacionais, A inserção de um personagem é capaz de aumentar o engajamento do aluno pelo tema a ser aprendido, não apenas em conteúdos que envolvam a tecnologia, como programação, mas também nas disciplinas pertencentes ao currículo base das escolas, como língua estrangeira, geografia, história e matemática (Carlos Germano, 2020). O uso de robôs humanoides, capazes de demonstrar emoções e responder a estímulos externos tornam as suas aplicações muito mais atrativas para crianças e adolescentes.

Estudos de caso recentes verificaram que o uso de robôs sociais responsivos aumenta o engajamento de crianças do espectro autista nas relações sociais. David Silvera-Tawil e Scott Andrew Brown (2019) utilizaram o robô social KASPAR (Figura 1.1) como ferramenta para auxiliar na terapia de crianças do espectro autista durante um período de cinco semanas, com a finalidade de ajudá-las a melhorar habilidades de comunicação, foco e atenção. Durante o período de intervenção foram feitos diversos acompanhamentos e percebeu-se que KASPAR é sim capaz de aumentar o engajamento da criança durante a terapia e diminuir os comportamentos de enfrentamento, fornecendo um ambiente onde novas habilidades podem ser introduzidas e praticadas com um terapeuta como mediador.

Figura 1.1 – Kaspar sendo controlado por um pesquisador.

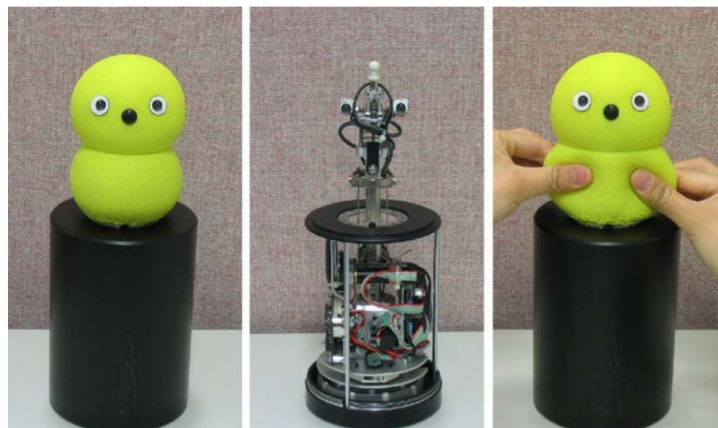


Fonte: Adaptado de Silvera-Tawil e Brown (2019)

Outro exemplo que pode ser citado é o desenvolvimento do robô Keepon (Figura

1.2), um pequeno robô desenvolvido a fim de facilitar a interação com humanos, principalmente bebês e crianças. Em um ambiente controlado (KOZIMA; MICHALOWSKI; NAKAGAWA, 2009), as crianças que interagiram com o Keepon demonstraram curiosidade com relação ao robô, entrando, gradativamente, em uma interação exploratória e social. Apesar de completamente diferente de um humano o Keepon possui a capacidade de expressar emoções simples que, combinada com a suavidade natural da sua construção, permite que crianças e bebês compreendam intuitivamente a carga emocional das interações.

Figura 1.2 – Robô Keepon e sua constituição.



Fonte: Adaptado de Kozima, Michalowski e Nakagawa (2009)

Também é possível analisar os efeitos que robôs têm em pessoas mais velhas. O PARO (PARO Robots, 2014), robô terapêutico em formato de foca (Figura 1.3), foi desenvolvido para simular um animal de estimação, tendo como principal objetivo reduzir o stress de pacientes idosos com demência e depressão (Jacqueline Kovacs, 2014). O robô possui cinco tipos de sensores: sensores táteis, de luz, de audição, de temperatura e de postura, com os quais pode perceber as pessoas e seu ambiente. O processo do idoso cuidar do PARO estimula a interação entre pacientes e cuidadores e melhora o relaxamento e a motivação. Para especialistas o robô é capaz de gerar um elo para aspectos mais positivos da vida, ajudando as pessoas a reduzir os sentimentos negativos em prol de sentimentos e memórias positivas.



Figura 1.3 – PARO, o robô terapêutico.



Fonte: Adaptado de Jacqueline Kovacs (2014)

Na mesma linha de robôs de estimação, mas nesse caso para companhia e entretenimento, ao invés de terapia, um dos robôs mais famosos como companheiros são os robôs AIBO (SONY, 2020) desenvolvido pela SONY (Figura: 1.4). O AIBO é um robô em formato de cãozinho, programado para ter o comportamento de um animal de estimação, brincando com seus donos e interagindo com crianças e adultos. O AIBO é capaz de reconhecer pessoas, espaços, realizar truques e aprender coisas novas. Ele se tornou tão popular e querido por seus donos ao ponto que em alguns lugares do mundo surgiram empresas especializadas em realizar funerais (James Burch, 2020) para o pequeno robô quando não é possível mais fazer a sua manutenção pelo chamados "veterinários de robôs". É que durante um longo período a SONY parou a produção do pequeno robô, o que aumentou a demanda por este tipo de serviço pois modelos mais antigos ficaram sem a possibilidade de manutenção. Recentemente o robô voltou a ser produzido em nova versão.

Figura 1.4 – AIBO, o cão robô.

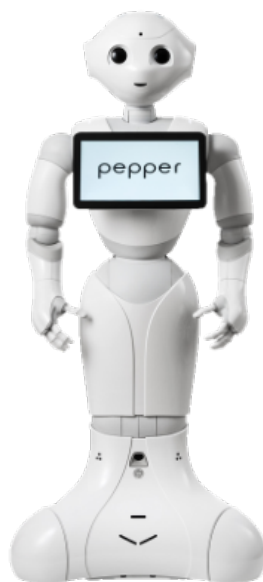


Fonte: Adaptado de SONY (2020)

Nas áreas comerciais o uso do robô Pepper (Figura: 1.5), desenvolvido pela Soft-

Bank Robotics tem sido cada vez mais difundido. O Pepper, ou "a" Pepper, já que às vezes é atribuído a esse robô um gênero feminino, é o um robô humanoide social capaz de reconhecer faces humanas e suas emoções básicas. Otimizado para interações com pessoas ele é capaz de se envolver em relações humanas não apenas através da tela estilo *tablet* em seu peito mas também por meio de conversações. O robô possui 20 graus de liberdade para movimentos naturais e expressivos. Ele possui reconhecimento de voz e é capaz de diálogos em diversos idiomas. Pepper também dispõe de sensores de toque, microfone, sensores infravermelhos, câmeras 2D e 3D e sonares para navegação autônoma. O equipamento já é usado como robô concierge em vários ambientes, como lojas e hotéis, aumentando a eficiência do processo administrativo, e em vários setores da área hospitalar, como assistente hospitalar, como comunicador ou gerador de dados em pesquisas científicas.

Figura 1.5 – Pepper.

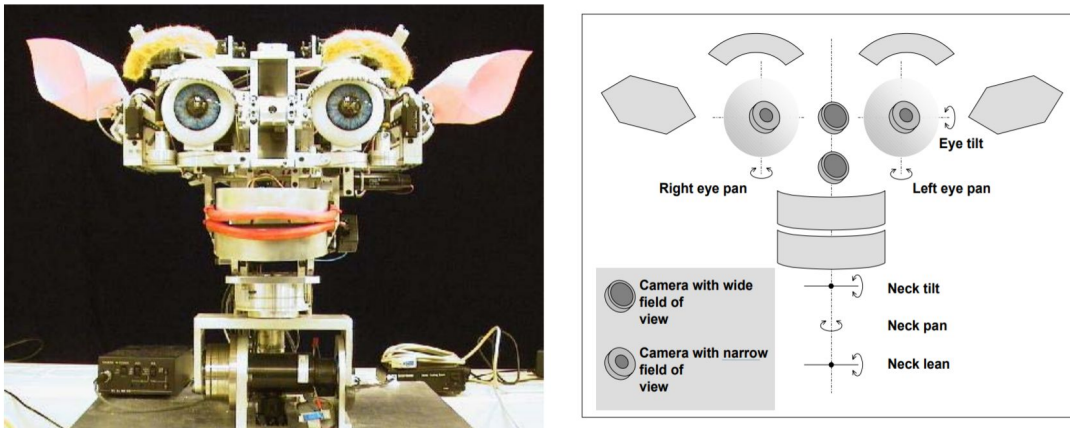


Fonte: Adaptado de SoftBank Robotics (2020)

Estudar sobre as capacidades de robôs em expressar emoções, para uma melhor interação social não é um tema recente: empresas e universidades vêm trabalhando há décadas a fim de desenvolver robôs e avatares que sejam capazes de se expressar. No trabalho *“Emotion and Sociable Humanoid Robots”*, a pesquisadora Dr. Cynthia Lynn Breazeal (BREAZEAL, 2002), pioneira na pesquisa das interações entre humanos e robôs, foca no papel da emoção e da expressividade na interação social entre humanos e robôs em diversos cenários. Para isso, ela apresenta os dados obtidos com as pesquisas utilizando o Kismet (Figura 1.6), robô desenvolvido pela mesma na década de 90 durante o

seu doutorado no laboratório de pesquisa do Instituto de Tecnologia de Massachusetts, o *MIT Media Lab*. O Kismet foi desenvolvido com capacidades expressivas, perceptuais e motoras adaptadas aos canais de comunicação humano, ou seja, ele é capaz de reconhecer e simular emoções, aumentando o engajamento das pessoas com a comunicação, e também promovendo o bem estar das pessoas que interagem com ele.

Figura 1.6 – Visão geral do Kismet e seus componentes.



Fonte: Adaptado de Breazeal (2002)

Como foi visto nos parágrafos acima, os serviços de assistentes, sejam eles robôs humanoides ou assistentes pessoais, são tecnologias extremamente flexíveis, estando cada vez mais presentes na vida das pessoas, seja na educação, saúde ou lazer. As suas funções podem ser reprogramadas, modificadas e otimizadas, e as aplicações são cada vez mais adaptadas e responsivas para o contexto na qual ela está inserida. Os robôs deixaram de ser a ficção da tela de cinema e se tornaram parte do cotidiano da sociedade. No restante deste capítulo serão definidos a justificativa e o escopo deste trabalho.

## 1.2 JUSTIFICATIVA

A comunicação é um processo muito importante em uma comunidade. Ao se comunicar, o humano, como ser social, é capaz de transmitir mensagens, intenções e emoções a um público, fazendo valer suas ideias e posicionamentos. Para essa comunicação ser bem sucedida é necessário que haja, ao menos, um emissor e um receptor. O emissor deve ser capaz de se expressar adequadamente, fazendo claras as suas ideias, motivações e propósitos, a fim de convencer, explicar ou gerar empatia com o receptor.

A forma como uma pessoa se movimenta e se expressa está intimamente ligada a sua personalidade e ao seu estado de espírito. Movimentos rápidos podem indicar excitação, alegria, nervosismo, enquanto movimentos mais lentos podem indicar letargia, calma,

tristeza. Além disso fatores ambientais e cognitivos produzem reações nas pessoas. Essas reações podem se manifestar de maneiras físicas, como sorrisos, diferentes formas de olhar, etc. O uso da linguagem corporal e da comunicação não-verbal vem sendo tema de diversos estudos, de forma a entender o papel de gestos e expressões faciais nas relações existentes entre os participantes do processo de comunicação.

Dessa forma, ao se introduzir um robô no processo de comunicação, tem-se a expectativa de que o mesmo seja capaz de manifestar, ou transmitir emoções. Para Breazeal (BREAZEAL, 2002) as expressões exercem um papel fundamental nas interações sociais entre humanos e robôs. Com robôs comunicadores não seria diferente, as expressões faciais, aliadas aos movimentos corporais podem definir intenções e emoções. Esse conjunto de fatores é essencial para que o robô tenha uma personalidade aos olhos do público.

Um estudo realizado por J. Yang et. al (JEONG; YANG; BALTES, 2020), analisou os resultados obtidos na competição "*The Humanoid Robot Magic Show*", O Show de Mágica do Robô Humanoide, em tradução livre, ocorrido no "*International Conference on Intelligent Robots and Systems 2018*" (IROS 2018). Nesta competição os robôs participantes devem performar shows de mágica. Tendo como base o primeiro lugar da competição, chegou-se à conclusão que a personalidade do robô é muito importante para manter a audiência e cativar o público em situações de apresentação.

A ideia de animar um elemento com o intuito de torná-lo crível para o público é extensamente utilizada em diversos estúdios de animações e jogos, de forma a criar figuras com personalidades cativantes e envolventes, gerando a ilusão de vida. Para Joseph Bates (BATES, 1994), em "*The Role of Emotion in Believable Agents*", personagens sem emoção são personagens sem vida, como uma máquina qualquer, portanto, a emoção, é o principal meio para que um personagem possa causar uma impressão de realismo.

A falta de expressão e credibilidade afeta, inclusive, o uso dos avatares de acessibilidade para LIBRAS. Muitos acabam não sendo bem aceitos dentro da comunidade surda devido à baixa capacidade deles em usar expressões não manuais, ou seja, expressões faciais e corporais que indiquem o objetivo, a valência e a intensidade daquilo que é dito através dos sinais. Para Brecailo (BRECAILO, 2012), a expressão corporal e facial é de extrema importância para a comunicação e mais ainda para comunicação em LIBRAS, trazendo clareza e objetividade ao sinal.

Dessa forma, exprimir emoções não se trata apenas de tornar uma apresentação mais bonita e agradável aos olhos do público, mas também de fazer valer as informações que são transmitidas, criando laços entre o comunicador e o telespectador. Com base nas justificativas acima, este trabalho tem como principal objetivo, melhorar a expressividade de um robô que se comunica com humanos, tornando seus gestos mais naturais e intuitivos com apresentações claras através de um gerador automático baseado em emoções.

### 1.3 ESTRUTURA DO TRABALHO

Este trabalho está organizado da seguinte forma:

**Revisão Bibliográfica:** É onde são apresentadas outras pesquisas relacionadas à expressão de emoções por robôs humanoides e a fundamentação teórica que norteia este trabalho. A fundamentação teórica está subdividida em três tópicos. O primeiro é o estudo da face e suas configurações para expressão de emoções. Em seguida serão apresentadas as formas de expressões corporais para representação de emoções, por fim serão expostos alguns estudos sobre a detecção de emoções em texto e suas aplicações.

**Materiais e Métodos:** Este capítulo está subdividido em quatro partes, a primeira é onde será feita a descrição do hardware a ser utilizado, a segunda será a apresentação do trabalho de detecção de emoções em texto. Em seguida será feita a exposição do algoritmo base de criação de movimento e como ele foi adaptado e otimizado para essa aplicação. Por fim será demonstrada a integração do trabalho de detecção de emoções com o algoritmo de geração de movimento.

**Resultados Alcançados:** Aqui serão apresentados os resultados obtidos após a implementação do algoritmo e da aplicação, bem como os resultados obtidos durante a aplicação de questionário para avaliação da capacidade do robô de expressar emoções.

**Conclusão:** Capítulo dedicado a discussões sobre possíveis melhorias no algoritmo e aplicações futuras.

### 1.4 OBJETIVOS

Neste capítulo vamos apresentar os objetivos principais deste trabalho.

#### 1.4.1 Objetivos Gerais

Tendo como base o robô Beo da empresa Qiron Robotics, este trabalho tem como objetivo principal o desenvolvimento de um algoritmo capaz de gerar apresentações orais, com movimentos e expressões dos olhos. Essas apresentações devem ter como base os sentimentos configurados manualmente pelo usuário ou detectados automaticamente, através de algoritmo treinado com redes neurais artificiais, em um texto. Esse gerador deve ser capaz de produzir uma série de combinações de movimentos, cenas, que, aliados ao texto dito pelo robô, sejam capazes de transmitir sentimentos e intenções ao espectador.

### **1.4.2 Objetivos Específicos**

- Criação e otimização do algoritmo gerador de cenas;
- Geração de um banco de dados com no mínimo 50 gestos para criação de movimentos;
- Pré-classificação de frases para correta seleção dos arquivos necessários para criação da apresentação;
- Representação gráfica da forma como as frases são classificadas;
- Geração de apresentações de textos, com voz, movimentação e expressão de olhos adequados;
- Aplicação de questionário para avaliação do estética dos banco de dados de gestos.

## 2 REVISÃO BIBLIOGRÁFICA

Neste capítulo iremos apresentar os trabalhos relacionados a expressão de emoções em robôs, bem como a fundamentação teórica envolvida para a detecção automática de emoções em textos escritos.

### 2.1 TRABALHOS RELACIONADOS:

Dentro dos grupos de pesquisa na UFSM, outro exemplo que pode ser citado é o desenvolvimento de um rosto para o robô DoRIS (CASTRO et al., 2019), um robô de uso doméstico. Esse projeto é o resultado de uma parceria entre o grupo de pesquisa Taura Bots da Universidade Federal de Santa Maria e o time de competição FURGBOT, da Universidade Federal do Rio Grande. O rosto de DoRIS pode ser observado na figura 2.1. Este rosto tem como objetivo reduzir o desconforto das interações entre humanos e robôs e também melhorá-las.

Figura 2.1 – Face do robô DoRIS



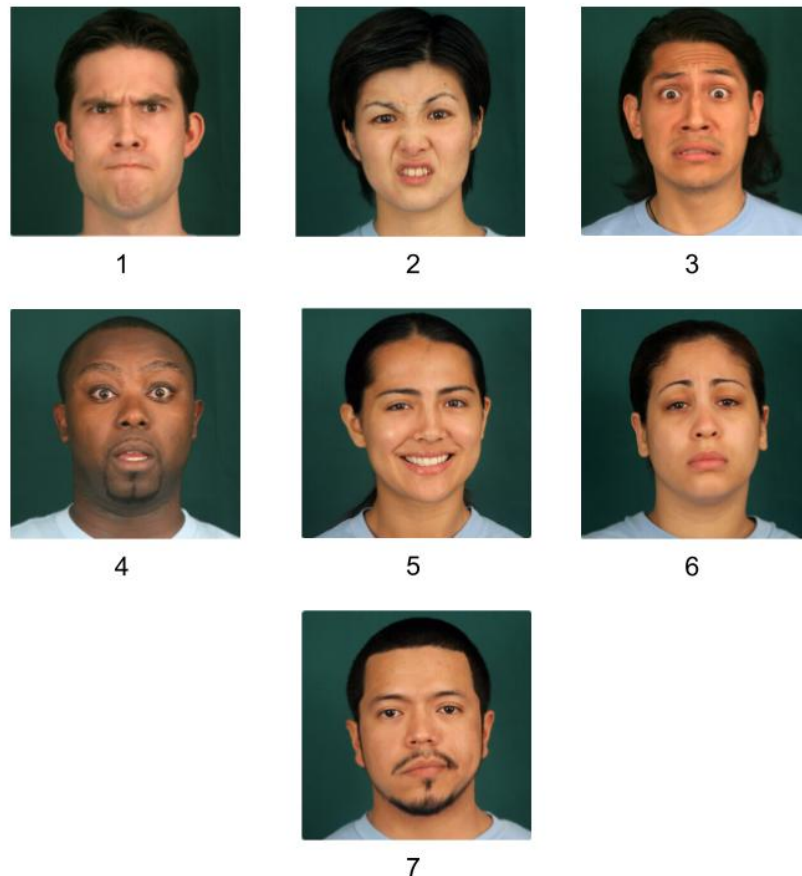
Fonte: Adaptado de Castro et al. (2019)

O rosto robô DoRIS lembra a face humana, com olhos grandes e expressão serena, semelhante a um *cartoon*, fazendo com que o robô seja mais atrativo e simpático aos olhos de quem vê. Para controle das diversas funcionalidades do rosto de DoRIS, são utilizados 12 servo-motores conectados a um microcontrolador, para que ela seja capaz de reagir socialmente ao ambiente, processando e devolvendo respostas adequadas, por meio de voz e expressões faciais.

O ser humano é capaz de expressar diversas emoções em diferentes níveis de intensidade. Para Paul Ekman (EKMAN, 1977), psicólogo americano, pioneiro no estudo

das emoções e como elas são representadas, sete são as emoções básicas que podem ser retratadas pela face humana. São elas raiva, nojo, medo, surpresa, felicidade, tristeza e desprezo, como retratadas na Figura 2.2 abaixo.

Figura 2.2 – Raiva (1), nojo (2), medo (3), surpresa (4), felicidade (5), tristeza (6) e desprezo (7)



Fonte: Adaptado de Paul Ekman (2020b)

Com seus estudos, Ekman (EKMAN, 1977) conseguiu obter evidências sobre a universalidade dessas emoções e suas características faciais, tais como posição das sobrancelhas, configuração de boca e grau abertura dos olhos. Ele também percebeu a existência de micro expressões, movimentos faciais que ocorrem em fração de segundos, involuntários e que expõem as verdadeiras expressões de uma pessoa quando esta pessoa tenta omitir ou fingir emoções.

A partir do trabalho de Ekman, podemos traçar as características faciais de cada emoção, como mostrado no Quadro 2.1 abaixo:



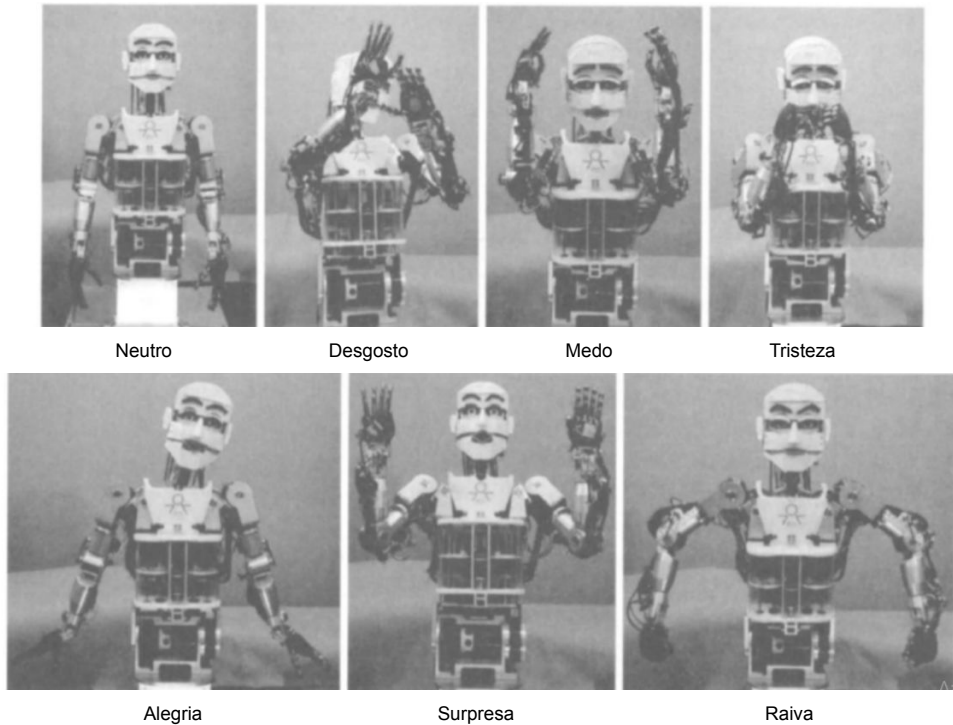
Quadro 2.1 – Características faciais de cada emoção

<b>Emoção</b>	<b>Olhos</b>	<b>Sobrancelhas</b>	<b>Boca</b>
<b>Raiva</b>	Abertos e olhando fixamente	Juntas e inclinadas para baixo	Lábios pressionados
<b>Nojo</b>	Enrugamento da pele na área mais próxima ao nariz	Inclinadas para baixo	Levemente aberta com o lábio superior ligeiramente saliente
<b>Medo</b>	Pálpebras inferiores tensas e pálpebras superiores erguidas	Juntas e erguidas	Lábios esticados em direção às orelhas
<b>Surpresa</b>	Pálpebra superior levantada e pálpebra inferior relaxada	Erguidas, porém não juntas	Mandíbula caída
<b>Felicidade</b>	Estreitos, com enrugamento da pele ao redor do olho	Relaxadas	Puxada para trás e mostrando os dentes
<b>Desprezo</b>	Abertos	Relaxadas	Canto da boca apertado e levantado
<b>Tristeza</b>	Olhar baixo	Parte interna das sobrancelhas levantadas juntas	Cantos da boca voltados para baixo

Fonte: Autora, a partir dos dados obtidos em Paul Ekman Group LLC (Paul Ekman, 2020b)

O trabalho de Kazuko Itoh et. al (ITOH et al., 2006) é um dos pioneiros utilizando robôs nessa temática. Ele analisou as formas como o robô WE-4RII pode representar diferentes emoções e suas intensidades, usando expressões faciais e movimentos corporais de braço, cabeça, ombros e tronco, como demonstrado na Figura 2.3. Em ambos os casos, do rosto DoRIS e do WE-4RII, pode-se notar certa similaridade entre os rostos dos robôs, tais como, a presença de sobrancelhas com controle de altura e inclinação, formas de mudar a direção do olhar e a abertura da boca. Isso permite que estes robôs manifestem diversas configurações faciais.

Figura 2.3 – Robô WE-4RII expressando diferentes emoções



Fonte: Adaptado de Itoh et al. (2006)

Através de questionários aplicados aos participantes da pesquisa realizada por Kazuko Itoh et al., demonstrou-se que a combinação correta de movimentos e expressões faciais pode transmitir emoções de forma adequada ao observador. Foram registradas as porcentagens de pessoas que reconheceram, adequadamente, as emoções transmitidas pelo WE-4RII em comparação às taxas obtidas na mesma pesquisa realizada com um modelo mais antigo do mesmo robô. Em ambos os casos houve uma alta taxa de reconhecimento da maioria das emoções. A emoção mais difícil de ser representada é a do medo, isso se comprova pela baixa taxa de reconhecimento obtida, cerca de 66% para o modelo mais novo. Em compensação as outras emoções possuem alta taxa de sucesso de discernimento, com taxas superiores a 88%.

Ekman também fez algumas definições sobre as maneiras como os movimentos fornecem informações. Em uma dessas definições se encontra a descrição dos chamados Ilustradores (Paul Ekman, 2020a), que são os movimentos que ocorrem durante a fala, usados para traçar o fluxo das ideias, enfatizar aquilo que é dito ou até demonstrar ações. Os ilustradores são elementos que variam de cultura para cultura.

Dessa forma, pode-se destacar a importância das características dinâmicas do movimento, as formas como ele pode variar, seja em altura, velocidade e amplitude dos movimentos. Movimentos rápidos e abruptos podem indicar raiva, movimentos ondulados e circulares podem indicar alegria, vibrações podem representar o medo, e assim por diante.

Porém, alguns robôs não são capazes de mudar suas feições faciais, fazendo com

que seja necessária uma movimentação corporal mais rica e variada para criar diferentes composições a fim de externar emoções. Esse é o caso do robô NAO, representado na Figura 2.4. Seus rosto possui forma definida e imutável, exceto pela mudança na cor de alguns LEDs RGB ao redor dos olhos.

Figura 2.4 – Robô NAO



Fonte: Adaptado de RobotLAB (2020)

O robô NAO é um dos robôs mais conhecidos atualmente, sendo mundialmente utilizado em pesquisas científicas, projetos de extensão e na área de entretenimento. Apesar de não ser capaz de expressões faciais, ele conta com uma série de servo motores, em seus braços, pernas e pescoço, que fazem com que ele seja capaz de se mover de forma muito parecida a uma pessoa real. Ele também pode manipular objetos, contando com a ajuda de câmeras, sensores infravermelhos, sensores de toque, acelerômetros e giroscópios, tudo isso para que ele seja responsivo e apto a ser utilizado em diversos ambientes. O robô NAO já é utilizado em programas educacionais, pesquisas científicas, indústrias e varejo devido ao grande número de possibilidades e projetos que podem ser desenvolvidos em sua plataforma.

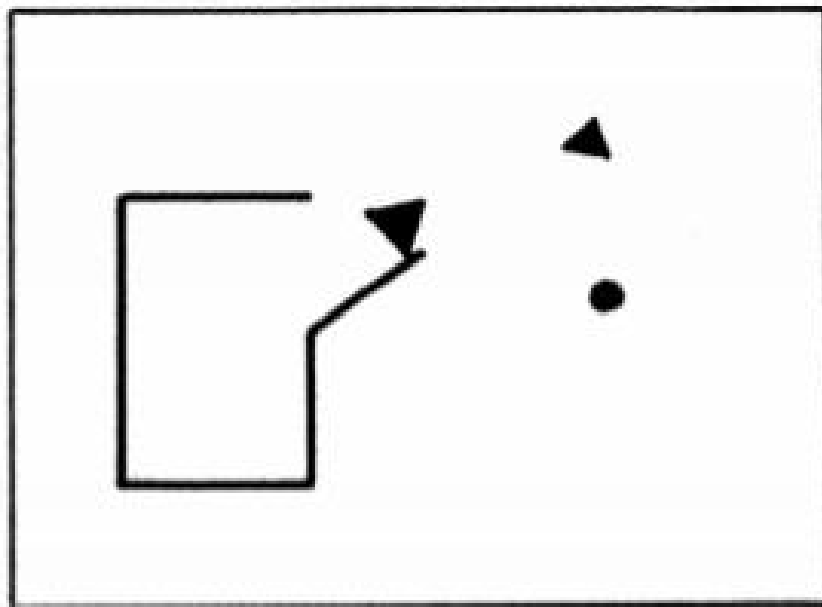
Aryel Beck et. al (BECK et al., 2010) estudou a percepção das emoções transmitidas pelo robô NAO através de poses, usando um algoritmo que combina (interpola) entre um conjunto definido de poses principais para gerar novas poses que compõem o movimento como um todo. Essa combinação é feita a fim de criar uma zona de afetividade e por fim melhorar a manifestação de emoções nos robôs.

Os pesquisadores chegaram à conclusão que a ausência de expressões faciais não é uma barreira para a demonstração de emoções por parte do robô, tendo em vista que a maioria dos participantes da pesquisa conseguiu reconhecer a emoção principal de cada

pose. Então, não apenas ser capaz de gerar um movimento, mas sim um movimento que possa ser interpretado, é um fator essencial na síntese de gestos a serem reproduzidos por um avatar, ou um robô.

De fato, emoções podem ser expressadas até mesmo na ausência de qualquer tipo de antropomorfismo. Um estudo experimental, feito por Fritz Heider e Maianne Simmel<sup>1</sup> (HEIDER; SIMMEL, 1944), apresentou a diversos grupos de pessoas um pequeno vídeo com movimentos de três figuras geométricas Figura 2.5. No vídeo as figuras interagiam entre si de diversas maneiras diferentes e, no final, foi pedido que os grupos fizessem descrições daquilo que viram, a maioria dos participantes relataram histórias, como se aquelas formas geométricas fossem personagens animados com personalidades.

Figura 2.5 – Figuras apresentadas por Fritz Heider e Maianne Simmel



Fonte: Adaptado de Heider e Simmel (1944)

Independente das histórias descritas pelos participantes, cada forma geométrica foi humanizada, sendo elas relatadas como tipos diferentes de pessoas, com sentimentos e motivações, que variam de acordo com a movimentação e aparência de cada uma. Dessa maneira a cena representada pelos triângulos e pelo círculo, de alguma forma evoca sentimentos nos observadores, que acabam influenciando na interpretação do vídeo como um todo, ainda que as formas em si não tenham nenhuma componente antropomórfica. A percepção da personalidade de cada forma foi feita exclusivamente através dos movimentos.

As características das personalidades, apresentadas pelos participantes do estudo, para o triângulo maior são as de agressividade, temperamento irritado, vilão, pessoa possessiva e que tira vantagem do tamanho em detrimento dos menores. Para o triângulo

<sup>1</sup>Disponível em: <https://youtu.be/n9TWwG4SFWQ>

menor, as principais qualidades são heroico, valente, corajoso e astuto. O círculo é descrito como um personagem fraco, tímido e dependente. Durante a pesquisa a maioria das particularidades de cada uma das formas geométricas foram atribuídas ao seu comportamento, ou seja, a forma como se movimentam durante o vídeo.

Neste trabalho buscamos expressar emoções utilizando a movimentação corporal através da variação na velocidade, altura e amplitude dos movimentos. Eles devem adequados a cada parâmetro das emoções que se quer representar. A velocidade do movimento será definida pelo nível de excitação, ao mesmo tempo que a altura e amplitude serão definidos pela valência da emoção, quanto mais alta a valência, alto e amplo será o movimento, podendo haver algumas variações por questões estéticas e funcionais

Outro desafio é aliar esse gestual a um conteúdo falado, de forma a representar aquilo que é dito. No artigo *“Style-Controllable Speech-Driven Gesture Synthesis Using Normalising Flows”*, os autores Alexanderson et al. (2020) propuseram um sistema baseado em redes neurais para sintetizar gestos a partir de uma entrada de voz. Eles conseguiram, de forma convincente, performar gestos em agentes virtuais renderizados em ambiente tridimensional usando dados de voz com o corpo humano inteiro, saindo do foco de partes isoladas do corpo, como mostrado na Figura 2.6 abaixo.

Figura 2.6 – Movimentos gerados por entrada de voz



Fonte: Adaptado de Alexanderson et al. (2020)

Neste trabalho os autores conseguiram, com sucesso, habilitar uma geração contínua de gestos naturais e estáveis, através do treinamento de redes neurais, com controle opcional. Porém, eles também observaram que existe uma complexidade em fornecer o modelo certo para essa geração. Modelos mal configurados apresentaram comportamento indesejáveis, ficando presos em poses não naturais, ou não acompanham a cadência da fala.

Dessa forma, a fim de aliar aquilo que é dito com a movimentação do robô, se gera a necessidade de analisar o conteúdo textual a ser reproduzido. A análise de textos através de algoritmos computacionais já é algo estudado com a finalidade de melhorar a qualidade de vida das pessoas. No trabalho de Poulin et al. (2014) a análise de notas clínicas retiradas da amostra nacional de registros médicos da Administração de Veteranos dos EUA, os autores conseguiram construir um modelo de predição ao suicídio com precisão de 65% ou mais. Para isso eles utilizaram um algoritmo de aprendizado baseado em algoritmos

genéticos. Os resultados permitiram que médicos selecionassem pacientes psiquiátricos, aparentemente saudáveis, para uma avaliação contínua dos riscos de suicídio.

No âmbito empresarial o uso do reconhecimento de emoções e opiniões em texto também podem ser muito úteis para melhorar o relacionamento das empresas com seus clientes. A análise de textos contidos em *e-mails* de reclamação de clientes, (GUPDA; GILBERT; FABBRIZIO, 2012) pode identificar características importantes para a categorização das mensagens. Através disso as empresas podem diferenciar, de forma automática, mensagens que contenham emoções negativas e opiniões negativas, ou até mesmo, detectar a reportar possíveis ameaças.

Entretanto a análise de das emoções em textos precisa levar em consideração o contexto na qual ela está inserida (MEDHAT; HASSAN; KORASHY, 2014), levando em consideração onde e por quem o texto é digitado. Atualmente, as maiores fontes de dados para as pesquisas em torno desta área são microblogs, blogs, fóruns da internet e redes sociais, como Twitter, onde pessoas expressam opiniões e sentimentos em pequenos textos de até 280 caracteres popularmente conhecidos como *tweets*.

O Twitter oferece uma API pública, que permite a análise de *tweets* publicados à medida que eles ocorrem, o "*We Feel*". O "*We Feel*" é um sistema de análise em tempo real de variações globais e regionais que analisa o conteúdo emocional de *tweets*. Através dele pesquisadores (LARSEN et al., 2015) analisaram e mapearam certas emoções e correlacionaram elas com padrões de postagens semanais e geográficos. Apesar de algumas limitações, cerca de 22000 textos por minuto foram analisados ao longo de 12 semanas, os pesquisadores observaram relações entre as emoções detectadas nos *tweets* e os índices de ansiedade e suicídios fornecidos pela Organização Mundial da Saúde.

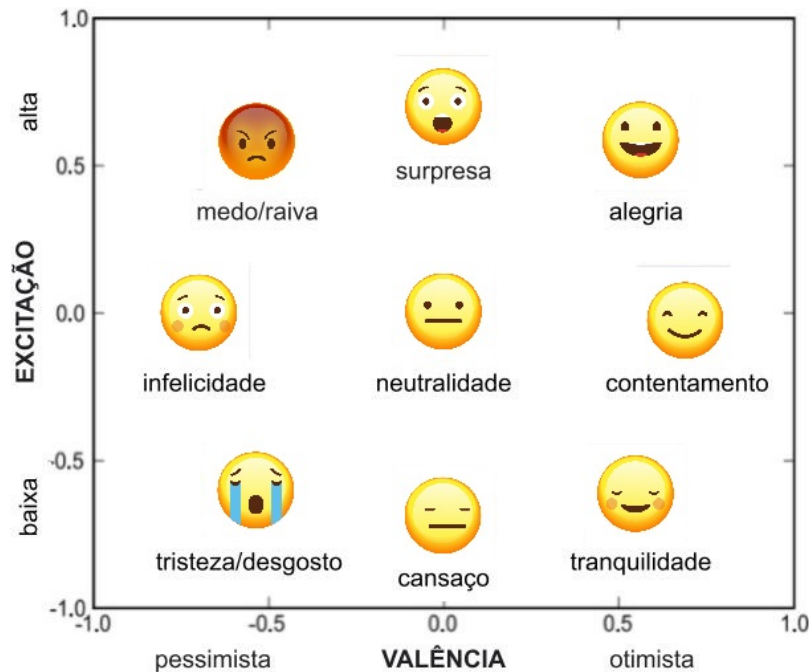
A detecção de emoções em linguística computacional é o processo de identificar emoções discretas expressas em texto (SEYEDITABARI; TABARI; ZADROZNY, 2018). Esse processo possui diversas aplicações, incluindo áreas empresariais, *de marketing*, e tratamentos médicos. As informações obtidas, quando há grande quantidade de dados envolvidos, podem mudar as formas como empresas e organizações oferecem seus serviços e tomam as decisões de suas políticas internas e externas.

Todo o texto possui informações implícitas em seu conteúdo. A escolha de palavras, bem como a combinação delas, são capazes de transmitir sentimentos, opiniões e contextos sobre as informações que são transmitidas. A forma como o texto é escrito é de extrema importância para que aquele que o lê seja capaz de interpretá-lo. Ao automatizar parcialmente essa interpretação, é vital que o software seja capaz de perceber as relações entre as palavras e classificar as emoções latentes nelas.

Transpondo esse conhecimento para os textos escritos, estima-se que eles carreguem os mesmos nuances emocionais que os estudados por Ekman. Na tentativa de mapear as diversas emoções dentro de um espaço de apenas duas dimensões, Du et al. (2007), parametrizam as emoções em um espaço 2D (Figura 2.7). Pode-se observar a

presença, não apenas das emoções estudadas por Ekman, mas também outros estados de espírito, tais como a neutralidade, o cansaço e a calma.

Figura 2.7 – Emoções representadas em um plano 2-D



Fonte: Adaptado de Du et al. (2007)

O gráfico da Figura 2.7 representa, no eixo horizontal a valência da emoção, ou seja, o quão positiva ou negativa uma emoção pode ser. Já o eixo vertical representa o nível de excitação da emoção. Essas dimensões ajudam a definir certos parâmetros, como altura e velocidade dos movimentos para cada uma das emoções. Por exemplo movimentos com alta excitação, como raiva e alegria serão movimentos mais rápidos e acelerados.

A presença de outras emoções, pode ser explicada pelas teorias e estudos de James A. Russell e Albert Mehrabian (RUSSEL; MEHRABIAN, 1977), que propõem uma classificação das emoções de acordo com três dimensões independentes e bipolares, são elas: valência, excitação e controle, mais conhecidas como 3-D VAC. Neste estudo os autores tabelaram os valores VAC de 151 estados de espírito.

A detecção das emoções em textos são tipicamente feitas de três maneiras (BINALI; WU; POTDAR, 2010). A primeira é baseada em palavras chaves, a segunda baseada em aprendizado e, por último, uma forma híbrida das duas primeiras. Dessa forma, a análise das palavras em um sistema de aprendizado, da maneira como elas estão organizadas e seus sinônimos, pode fornecer dados significativos dos contextos nos quais as elas estão inseridas.

### 3 MATERIAIS E MÉTODOS

Neste capítulo iremos descrever os métodos utilizados para o desenvolvimento deste trabalho. Parte deste projeto foi produzido em parceria com o estudante de Engenharia de Computação da Universidade Federal de Santa Maria, Moisés Goulart de Oliveira, que forneceu a sua pesquisa para detecção da emoção em texto de forma que pudesse ser utilizada para selecionar a emoção a ser representada pelo robô.

Em seguida faremos a descrição de *hardware* e da composição estética do movimento. Por fim será discorrido sobre o algoritmo de geração movimento, e a integração do algoritmo com a detecção da emoção nos textos.

É importante ressaltar que a metodologia escolhida para este trabalho é mista, fazendo abordagens tanto qualitativas quanto quantitativas para a obtenção dos resultados.

#### 3.1 DETECÇÃO DA EMOÇÃO EM TEXTO

A partir da parceria com o aluno Moisés Goulart de Oliveira, foi possível selecionar uma técnica de detecção de emoções em texto capaz de usar vetores de baixa dimensão que representam as palavras, esses vetores são conhecidos como *word embeddings*. Esses vetores são capazes de melhorar a resposta dos classificadores de texto em tarefas de processamento de linguagem natural (GOLDBERG, 2017). Em seu trabalho de conclusão de curso, em andamento, (OLIVEIRA, 2021) o estudante tem como principal objetivo analisar como as emoções em redes sociais, detectadas automaticamente via software, variam de acordo com acontecimentos recentes.

Para a detecção das seis emoções básicas de Ekman, foi feito o tratamento de um banco de dados de mais de cem mil sentenças curtas. Em seguida esse banco de dados foi usado para o treinamento da rede neural responsável pela detecção das emoções. A preparação dos dados foi feita em duas etapas principais. A primeira é a de tokenização, usado para representar as strings a serem processadas em pedaços menores, no formato *one-hot*. A segunda etapa é a de limpeza de dados, que retira itens não essenciais para a análise, como pontuações, links e nomes. Em seguida, os *one-hots* obtidos durante a preparação dos dados são transformados em *word embeddings*, ou seja, em vetores densos. Para cada palavra é obtido um vetor que serve de entrada para o modelo, que tenta inferir a emoção de cada frase. Em princípio, quanto maior for a quantidade de textos utilizados no modelo, melhor será o resultado final obtido.

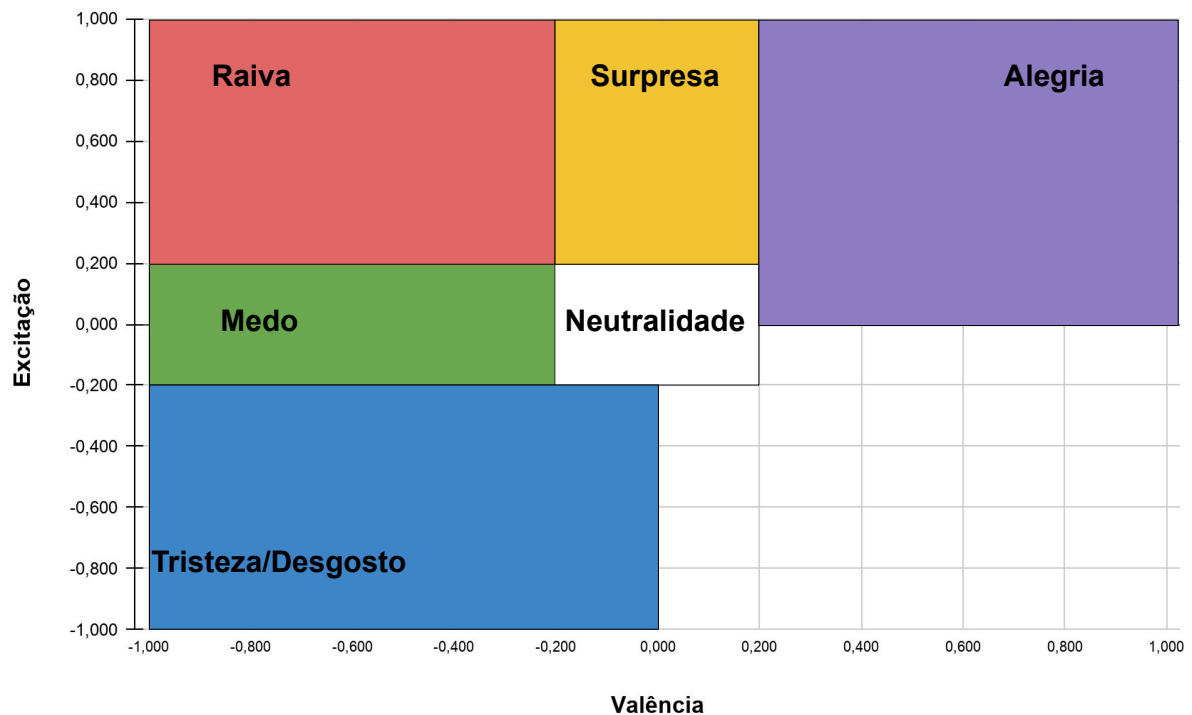
Após o treinamento da rede neural e verificação dos resultados o estudante concluiu que é possível utilizar esse método para detecção de emoções em texto. O resultado obtido para cada frase foi um vetor de 6 posições, onde cada uma das posições representa



uma emoção, raiva, medo, alegria, tristeza, surpresa e desgosto. Os valores deste vetor refletem a intensidade daquela respectiva emoção que foi detectada, podendo variar de 0 a 1.

Com base nesses estudos vimos a necessidade de mapear as emoções estudadas por Ekman, tomando em conta a suas valências e excitações. Portanto, considerando o estudo proposto por Du et al. (2007), ilustrado na Figura 2.7, localizamos as emoções estudadas por Ekman em um plano 2-D representado na Figura 3.1.

Figura 3.1 – Emoções de Ekman em um plano 2-D



Fonte: Autora

Se compararmos a Figura 3.1 com a Figura 2.7, algumas diferenças poderão ser notadas. Para simplificação do problema, tendo em vista a similaridade entre o sentimento de tristeza e o de desgosto, levaremos em consideração que ambas as emoções serão representadas apenas pela componente “tristeza” do vetor, o que faz com que o desgosto seja representado na mesma região do gráfico que a tristeza. Além disso, por não serem contempladas pelos estudos de Ekman, as emoções do canto inferior direito, contentamento, tranquilidade e cansaço, não serão representadas.

Levando-se em consideração o nível de valência e excitação de cada emoção é possível mapear o vetor emocional de 5 dimensões (onde cada componente representa a magnitude de cada respectiva emoção) em um plano 2-D, transformando-o em um vetor de apenas duas posições. Para isso adotamos um sistema de coordenadas polares, com centro na origem (região de “neutralidade”). Primeiramente determinamos a posição an-

gular de cada uma das emoções de acordo com o Quadro 3.1. Assim, definimos vetores polares de comprimento unitário, e ângulos conforme o Quadro 3.1, para representar cada uma das 5 emoções canônicas. Em seguida, modulamos cada vetor polar pela respectiva componente do vetor de 5 dimensões que representa a emoção da frase. Isso resulta em cinco vetores polares, onde cada um mantém a direção original, mas possui comprimento ajustado de acordo com a intensidade da respectiva emoção. Em seguida, os cinco vetores são transformados em coordenadas retangulares e somados, resultando nas componentes de Valência e Excitação.

Quadro 3.1 – Quadro com os ângulos de referência de cada emoção

<b>Emoção</b>	<b>Ângulo</b>
<b>Alegria</b>	45°
<b>Surpresa</b>	90°
<b>Raiva</b>	135°
<b>Medo</b>	180°
<b>Tristeza/Desgosto</b>	225°

Fonte: Autora

As coordenadas "x" e "y", de valência e de excitação, respectivamente, são úteis para entrada manual dessas variáveis para obtenção do movimento, ou seja, podemos obter a valência e a excitação a partir da emoção, ou, a emoção, a partir de uma entrada manual. Essa é uma possibilidade para aumentar as funcionalidades do algoritmo de geração de cenas.

### 3.2 DESCRIÇÃO DE HARDWARE

O robô utilizado para a execução deste trabalho é o robô Beo (Figura 3.2), produzido pela Qiron Robotics, empresa brasileira sediada na incubadora tecnológica da Universidade Federal de Santa Maria, a ITSM.

Com robôs sociais e carismáticos a Qiron gera engajamento para seus clientes, expondo suas marcas de maneira divertida e atrativa para o público geral. O modelo de robô Beo permite diversas interações com público, através de comandos via aplicativo *mobile*, controle por voz, reconhecimento facial e apresentações programadas.

Figura 3.2 – O robô Beo



Fonte: Adaptado de Qiron Robotics (2020)

A face do robô ainda conta com dois *displays* OLED azuis de 0.96 polegadas, com protocolo de comunicação I2C, que proporcionam a visualização dos desenhos dos olhos do robô de forma nítida e clara, permitindo a mudança que representa cada emoção.

Por se tratar de um robô humanoide, o robô Beo é feito com o intuito de interagir com pessoas e foi planejado de forma a gerar empatia e atrair a atenção. Suas formas arredondadas convidam o usuário para se aproximar. Com cerca de 43 centímetros de altura esse robô conta com braços móveis, cada um com 3 graus de liberdade e uma cabeça, também com 3 graus de liberdade, como mostrado na Figura 3.3.

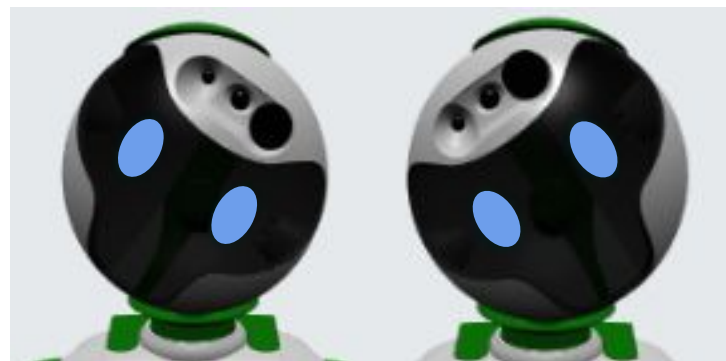
Figura 3.3 – Graus de liberdade da movimentação da cabeça do Beo



Guinada (Yaw/Paw)



Arfagem (Pitch/Tilt)



Rolagem (Roll)

Fonte: Autora

Para a movimentação dos braços e da cabeça são usados servo-motores Dynamixel XL-320. Este é um modelo de servo-motor potente para seu tamanho, muito utilizado na robótica, caracterizado pela sua leveza e pequeno tamanho (mostrado na Figura 3.4).

Figura 3.4 – Servo-atuador XL-320



Fonte: Adaptado de Robotis (2020)

Produzido com estrutura em plástico, este servo-motor é capaz de retornar ao dispositivo controlador diversas informações, tais como posição atual, temperatura e tensão. Segue, no Quadro 3.2, as especificações técnicas deste atuador.

Quadro 3.2 – Características do Servo Motor XL-320

Item	Especificação
Taxa de Comunicação	7343 bps 1 Mbps
Resolução	0.29°
Alcance	0° a 300°
Peso	16.7 [g]
Dimensões	24 x 36 x 27 [mm]
Torque	0.39Nm @ 7.4V, 1.1A
Tensão de alimentação	6V até 8.4V (Recomendado : 7.4V)
Protocolo de comunicação	Comunicação Serial Assíncrona <i>Half Duplex</i>
Conexão física	Barramento multiponto de nível TTL

Fonte: Adaptado de Robotis (2020)

Para processamento de algoritmos o Beo possui uma ASUS Tinker Board (Figura 3.5), um computador de placa única baseado na arquitetura ARM. Produzido em formato pequeno, ele é adequado ao tamanho do robô e oferece um bom desempenho e compatibilidade com todos os recursos utilizados durante as iterações do robô com o público. Também chamada de *Tinker*, essa placa possui 2 GB de memória de canal duplo LPDDR3, o que permite o desenvolvimento de diversas aplicações, incluindo o processamento de imagens e reprodução de mídias. Além disso a placa possui opções de conectividade padrão do fabricante, incluindo uma interface GPIO de 40 pinos, mantendo compatibilidade parcial com placas Raspberry Pi, conectividade de rede, *Wi-Fi* integrado e controlador *Bluetooth*. Além da memória interna a placa possui entrada para cartão de memória, o que aumenta sua capacidade de armazenamento.

Figura 3.5 – ASUS Tinker Board



Fonte: Adaptado de ASUS (2020)

Além disso, o Beo conta com diversos outros recursos muito úteis em suas aplicações, como bateria com autonomia de até duas horas e meia de pleno funcionamento, sistema de som integrado à base do robô, portas de conexões USB e de rede ethernet e WiFi. O robô também dispõe de uma câmera integrada com microfone, que permitem o reconhecimento de rostos e de fala.

### 3.3 COMPOSIÇÃO ESTÉTICA PARA REPRESENTAÇÃO DE EMOÇÕES

Em seus estudos Ekman descreve sete emoções: raiva, nojo, medo, surpresa, felicidade, tristeza e desprezo. Para simplificação do problema, neste trabalho foram escolhidas cinco dessas emoções, alegria, surpresa, raiva, medo e tristeza. Além disso, para melhor adaptação do usuário e por experiência da empresa Qiron Robotics, foi sugerida a inclusão da possibilidade da emoção neutra. Para cada emoção foram escolhidas as diferentes imagens que seriam reproduzidas nos *displays* do rosto a fim de estampar os olhos. As imagens escolhidas já faziam parte do banco de dados do robô, são descritos a baixo e ilustradas na Figura 3.6.

- **Alegria:** essa é uma emoção de valência positiva e seus olhos são caracterizados pela aparência estreita, remetendo a um sorriso;
- **Neutralidade:** aqui, não há a necessidade de expressão, portanto os olhos retratados são abertos, sem modificações;

- **Surpresa:** seus olhos são marcados por estarem bem abertos, portanto essa emoção será mais marcada pelo movimento de braços e cabeça que pela expressão dos olhos;
- **Raiva:** com muita ênfase, em humanos, na posição das sobrancelhas, os olhos da raiva são representados com imagens que emulam a curvatura da sobrancelha inclinada para baixo;
- **Tristeza:** a tristeza, assim como o medo, é uma emoção negativa, possui sobrancelhas unidas com curvatura na sua parte superior externa;
- **Medo:** os olhos que representam o medo possuem curvatura nas áreas superiores externas, simulando as sobrancelhas unidas em pessoas.

Figura 3.6 – Olhos para as expressões de alegria, neutralidade, surpresa, raiva, tristeza e medo



Alegria



Neutralidade/Surpresa



Raiva



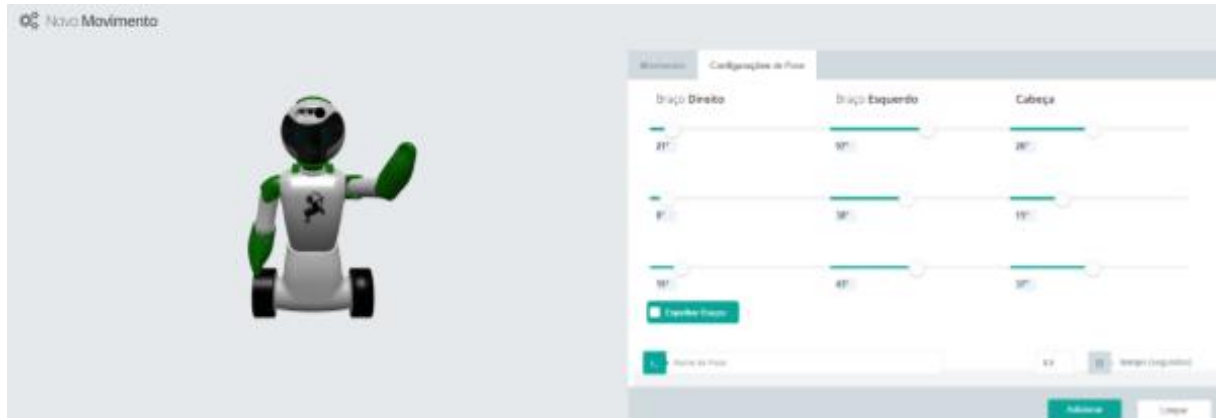
Tristeza/Medo

Fonte: Autora

O banco de dados dos gestos utilizados neste trabalho leva em consideração a

altura, velocidade e amplitude dos movimentos. Cada gesto foi estruturado por uma série de poses de referência, que geram poses intermediárias por interpolação, variando de uma pose de referência para outra, em uma determinada velocidade. Para essa composição foram usados os sistemas de criação de gestos da Qiron Robotics, Figura 3.7, que gera arquivos de extensão .mbeo, que podem ser lidos e executados pelo robô. Para cada gesto criado também foi feito um vídeo para avaliação.

Figura 3.7 – Ambiente para criação de Movimentos



Fonte: Autora

Cada conjunto de gestos é criado tendo como base o nível de excitação e valência de cada uma das emoções que serão representadas. Em termos gerais, a valência define a altura e amplitude dos gestos. Quanto mais positiva a valência mais alta e ampla é a composição. Quanto mais negativa a valência mais baixa e menos ampla será a composição. Já o nível de excitação define a velocidade da interpolação das poses. Gestos de maior excitação são mais rápidos, enquanto gestos de menor excitação são mais lentos.

Foram criados um total de 66 gestos, sendo eles 14 de alegria, 14 de surpresa, 11 de raiva, 13 de medo, 7 de tristeza e 7 de neutralidade. Após análises, foram descartados 24 gestos, sendo eles 5 de alegria, 6 de surpresa 3 de raiva, 8 de medo e 2 de tristeza. O critério de descarte foi a baixa ênfase a emoção que os movimentos deveriam representar, conforme julgamento da autora. Portanto, o banco de dados conta, atualmente, com 42 arquivos<sup>1</sup>. Para cada emoção os seguintes critérios foram estabelecidos de acordo com análises de apresentações anteriores feitas pelo robô para clientes da empresa.

- **Alegria:** emoção de valência positiva e alto nível de excitação, portanto foram criados gestos com altura geral maior, com braços acima da cabeça e com sentido ascendente. As poses desse movimento possuem alta velocidade e tem como objetivo simular gestos que remetem a comemorações (Figura 3.8);

<sup>1</sup>Disponível em: <https://youtu.be/t88yTzCzKGY>



Figura 3.8 – Alegria: quadro a quadro



Fonte: Autora

- **Surpresa:** possui valência neutra, com alto nível de excitação, suas poses começam de forma lenta e com altura de braço mediana, após algum tempo o movimento de braços acelera de forma repentina, simulando um gesto de susto (Figura 3.9);

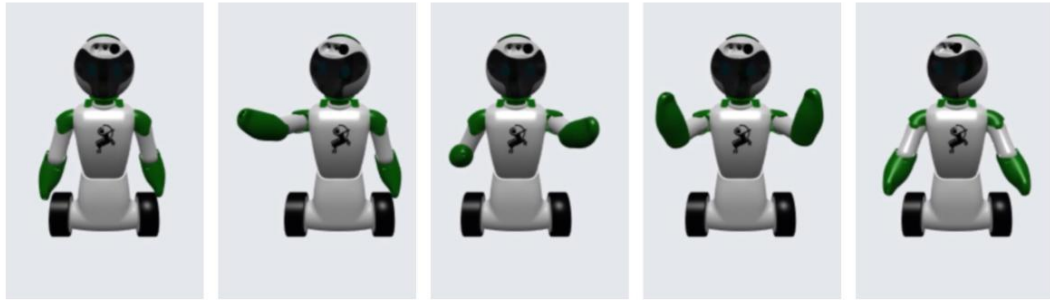
Figura 3.9 – Surpresa: quadro a quadro



Fonte: Autora

- **Raiva:** a raiva possui valência negativa e grande excitação, os movimentos dos braços são rápidos e de altura mediana, a cabeça faz movimentos de negação, oscilando horizontalmente em torno do eixo do pescoço, demonstrando contrariedade (Figura 3.10);

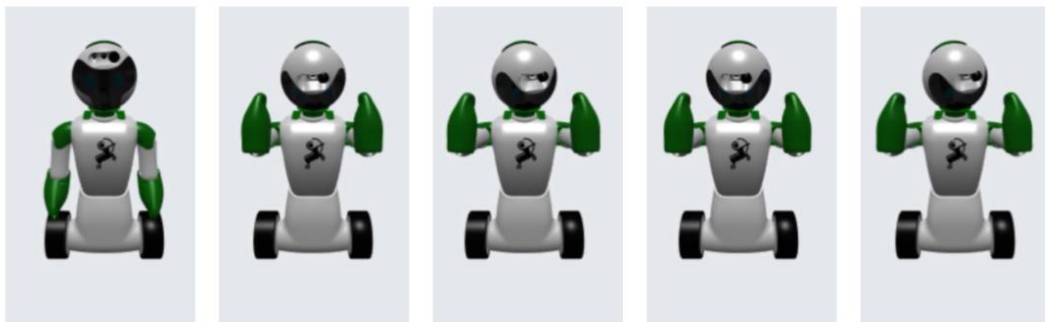
Figura 3.10 – Raiva: quadro a quadro



Fonte: Autora

- **Medo:** o movimento possui valência negativa, porém excitação mediana, portanto os movimentos de braços são curtos, vibrando, com objetivo de simular uma situação de estremeimento, a cabeça faz movimentos para os lados (Figura 3.11);

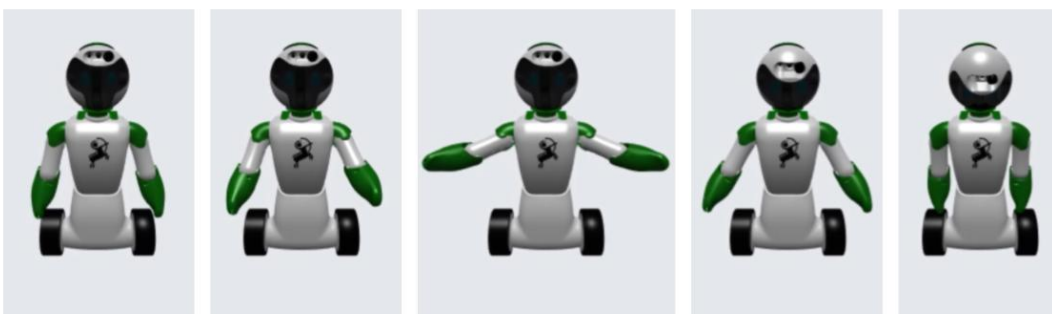
Figura 3.11 – Medo: quadro a quadro



Fonte: Autora

- **Tristeza:** emoção de baixa excitação e valência negativa, portanto os gestos são mais lentos e com altura de braço abaixo do nível do pescoço, em sentido descendente, a cabeça faz movimentos de negação (Figura 3.12);

Figura 3.12 – Tristeza: quadro a quadro



Fonte: Autora

- **Neutro:** valência neutra e excitação mediana, os movimentos de braço são centralizados e, algumas vezes, espelhados. Cabeça centralizada e com pouca movimentação, velocidade mediana entre as poses (Figura 3.13).

Figura 3.13 – Neutro



Fonte: Autora

O banco de dados é organizado em pastas, dessa forma o algoritmo que gera a apresentação final pode acessar os arquivos com maior facilidade, e também permite a fácil expansão e edição do banco de dados.

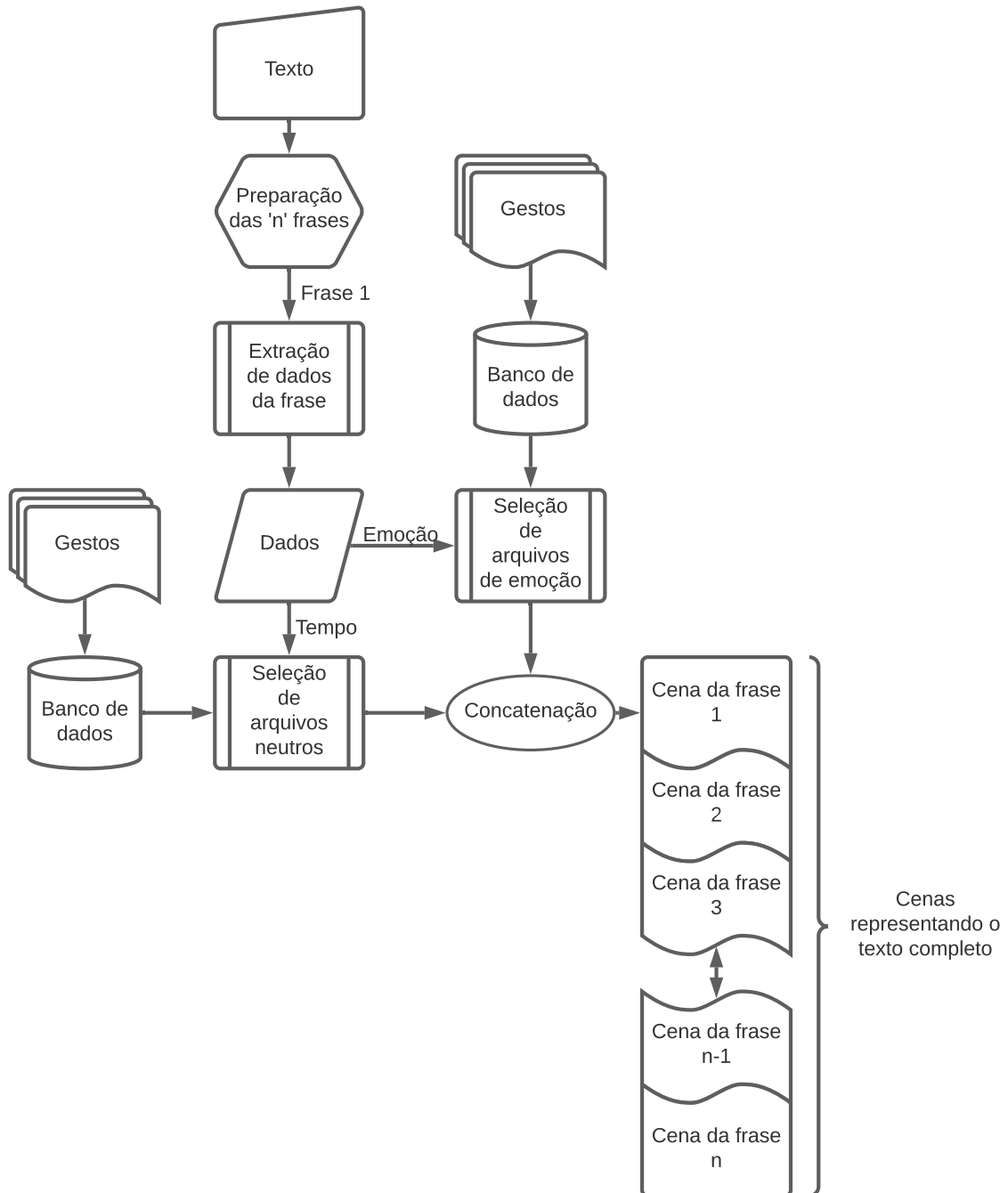
### 3.4 DESCRIÇÃO DO ALGORITMO

O principal objetivo deste trabalho é produzir um gerador de apresentações que seja capaz de gerar uma apresentação completa, com voz, expressão dos olhos e movimento, tendo como base as emoções automaticamente detectadas no texto, através dos *word embeddings*.

Dessa forma, ao se inserir um texto no algoritmo, ele é processado de forma a dividi-lo em frases menores. Nessas frases é aplicada a técnica de detecção de emoções em textos. Em seguida o algoritmo compila os arquivos criados no banco de dados de forma a combiná-los para criar uma sequência de gestos (cena), cujo tempo de duração seja dado pelo tempo das frases extraídas do texto, enquanto as características estéticas do movimento, altura, velocidade e amplitude, são dadas pela emoção extraída automaticamente das frases.

Na Figura 3.14, é apresentada uma versão simplificada do algoritmo que irá ser executado. A sequência lógica é pensada de forma a variar o tamanho da cena e das apresentações de acordo com o tempo que o robô leva para vocalizar todo o texto. O cálculo do tempo de fala se dá através do tamanho do arquivo de áudio gerado.

Figura 3.14 – Fluxograma do algoritmo para detecção automática da emoção.



Fonte: Autora

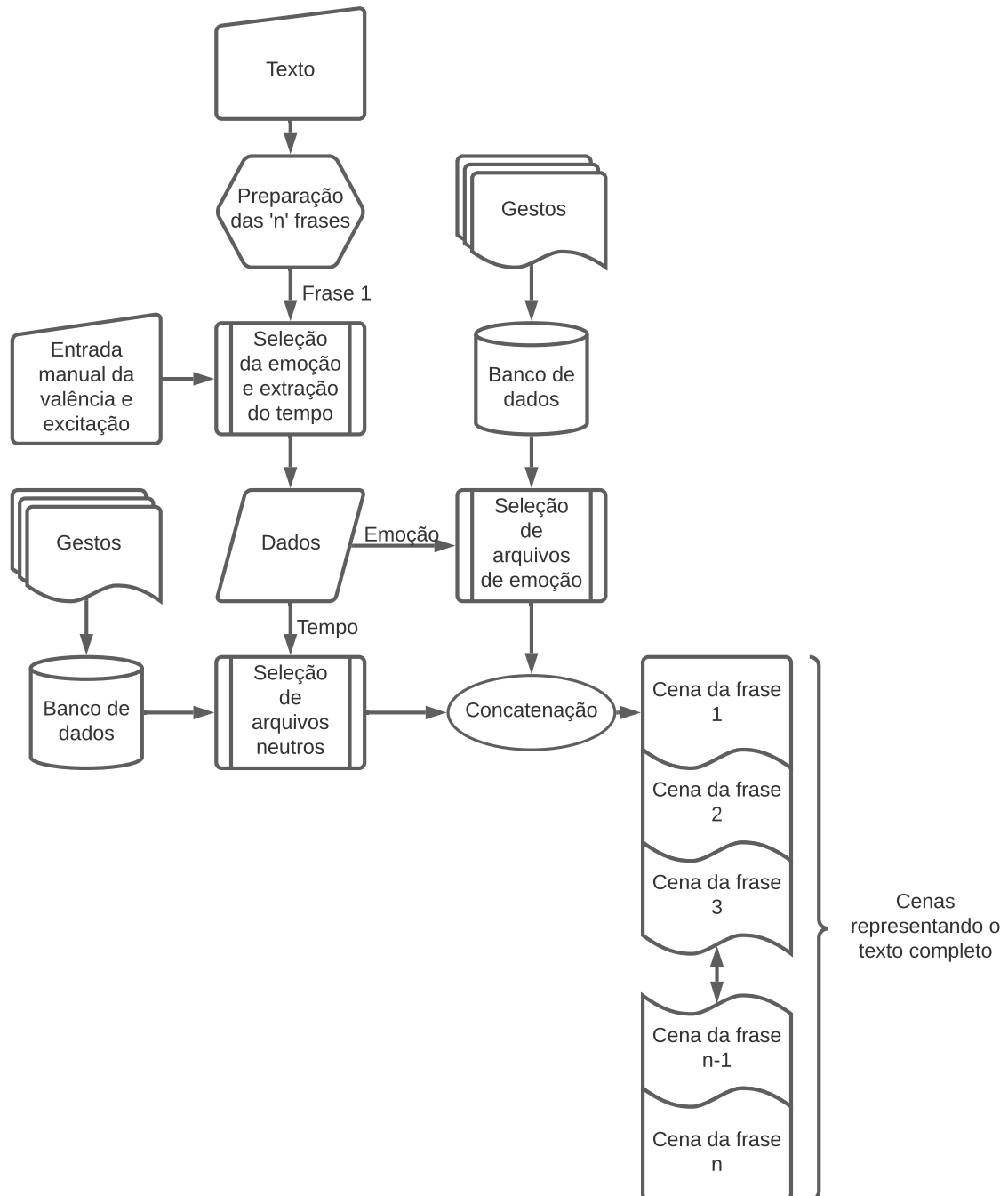
Após a seleção da emoção a ser representada o algoritmo seleciona, aleatoriamente, do banco de dados arquivos de gestos neutros que, de acordo com o tamanho da frase, sejam agregados ao arquivo de emoção específica. Dessa forma os movimentos irão se somar até que a sequencia seja suficientemente grande para abarcar toda a duração da frase a ser representada.

Um texto, que seja dividido em "n" frases irá executar a lógica de geração do movi-

mento "n" vezes, uma vez para cada frase, lembrando que cada frase possui uma emoção automaticamente detectada associada a ela.

Também é possível fazer a inserção manual dos valores de valência e excitação. Neste caso o algoritmo então fica representado de acordo com o fluxograma da Figura 3.15.

Figura 3.15 – Fluxograma do algoritmo com inserção manual da valência e excitação.



Para a geração de cada cena o arquivo da emoção detectada é escolhido de forma aleatória, entre aqueles disponíveis para aquela emoção. Então este arquivo é processado e suas informações são armazenadas em vetores. São ao todo 10 vetores, um vetor para controle de posição de cada um dos nove servo-motores do robô, cujas componentes são as sequências de posições angulares que os respectivos servo-motores devem assumir ao longo do tempo, e um vetor para controle do tempo entre essas posições. Para os vetores dos motores, cada valor representa o ângulo do eixo de uma junta do robô, tendo como referência a posição zero, mostrada na Figura 3.16.

Figura 3.16 – Posição zero do robô



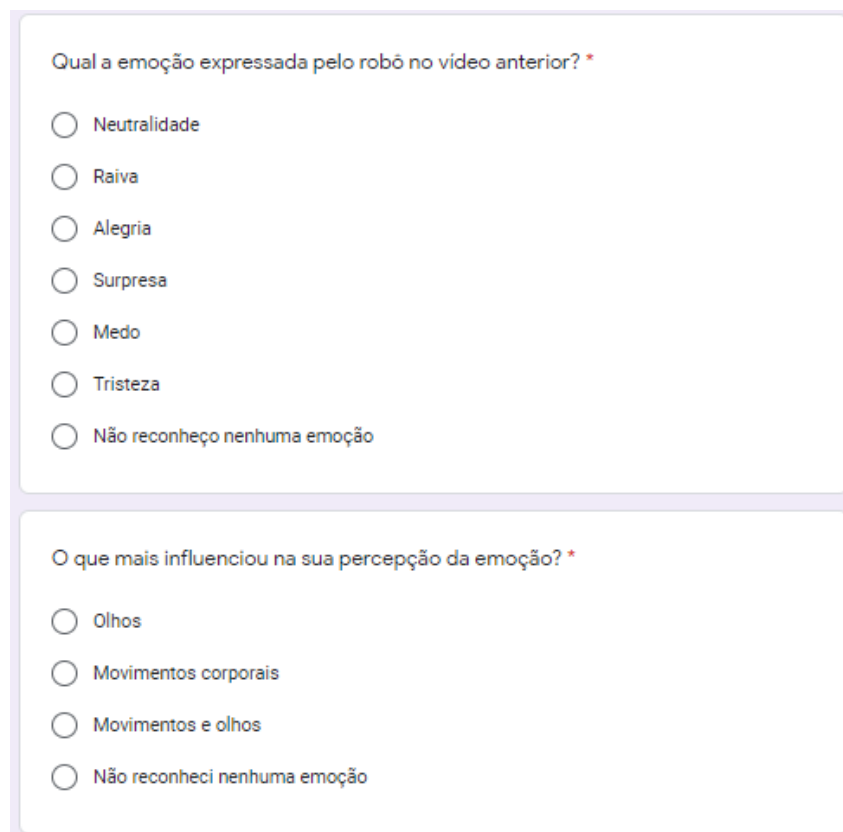
Fonte: Autora

Para completar a sequência são selecionados gestos neutros, cujos dados são processados para agrupar novas poses aos vetores de posição. O tempo total da cena deverá ser um pouco maior que o tempo total de fala da frase, dessa forma se garante que haverá vocalização de todo o texto.

## 4 QUESTIONÁRIO

Foi elaborado um questionário para a avaliação do banco de dados de gestos, bem como da representação dos olhos do robô para cada uma das emoções. Para esses questionários foram criados seis vídeos. Cada vídeo apresenta o robô encenando uma emoção, combinando movimentos e expressão dos olhos. Através de um formulário online, diversas pessoas foram convidadas a classificar as emoções demonstradas nos vídeos, de acordo com a sua percepção pessoal das cenas que eram apresentadas, como apresentado na Figura 4.1. Para a gravação dos vídeos foram utilizados apenas os gestos do banco de dados, as cenas não foram geradas pelo algoritmo.

Figura 4.1 – Apresentação do Questionário



Qual a emoção expressada pelo robô no vídeo anterior? \*

- Neutralidade
- Raiva
- Alegria
- Surpresa
- Medo
- Tristeza
- Não reconheço nenhuma emoção

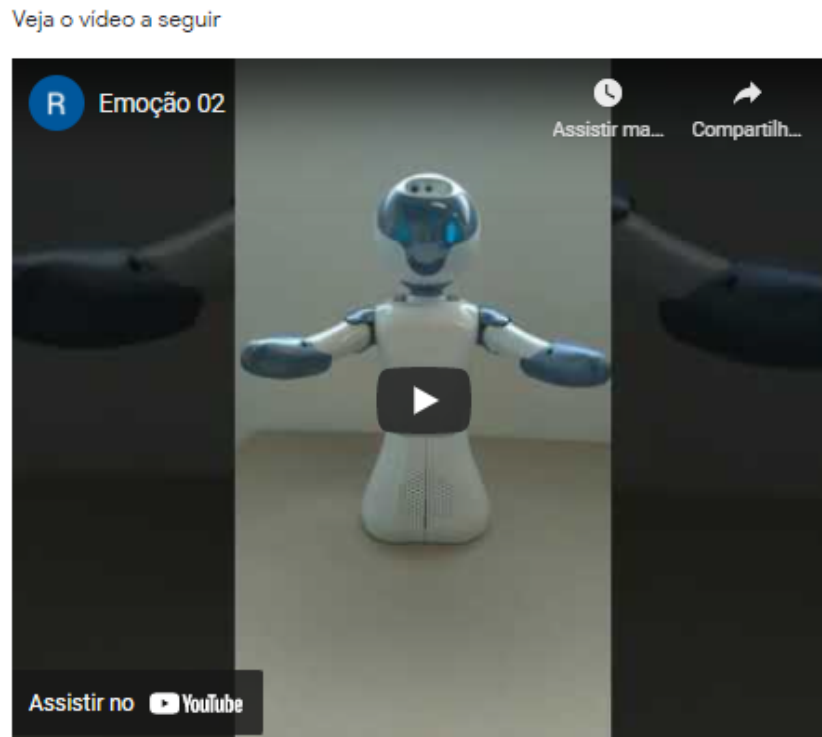
O que mais influenciou na sua percepção da emoção? \*

- Olhos
- Movimentos corporais
- Movimentos e olhos
- Não reconheci nenhuma emoção

Fonte: Autora

As cenas (Figura 4.2) apresentadas nos vídeos foram feitas utilizando movimentos neutros, combinados com gestos específicos de cada emoção.

Figura 4.2 – Apresentação do Questionário



Fonte: Autora

- Cena 01: Raiva, disponível em: <https://youtu.be/z3VUfmU4saE>;
- Cena 02: Alegria, disponível em: <https://youtu.be/u6SBq080Obg>;
- Cena 03: Medo, disponível em: <https://youtu.be/m71CpcFNtVc>;
- Cena 04: Neutro, disponível em: [https://youtu.be/lx\\_LfKBX81E](https://youtu.be/lx_LfKBX81E);
- Cena 05: Tristeza. disponível em: <https://youtu.be/LXO-p8clSZI>;
- Cena 06: Surpresa, disponível em: <https://youtu.be/lVua-d8Fciw>;

Foram aceitas respostas para o questionário durante quinze dias, também foram feitas divulgações em diversos grupos e redes sociais para aumentar o alcance. Ao final dos quinze dias foram obtidas 138 respostas cujos resultados serão apresentados nos próximos capítulos.



## 5 RESULTADOS

Neste capítulo apresentaremos os resultados obtidos, começando pelos resultados do uso do método para detecção da emoção em texto, seguido de algumas demonstrações estéticas de alguns textos utilizados pela empresa Qiron Robotics, incluindo o questionário aplicado à comunidade, e, por fim, os resultados obtidos na execução do algoritmo.

### 5.1 RESULTADO ESTÉTICO

O principal objetivo deste trabalho é a utilização do algoritmo para produção de apresentações que utilizem a correta expressão de emoções, em robôs, para a transmissão de mensagens a um público. Para a validação dos métodos utilizados foram produzidos alguns vídeos institucionais. O primeiro foi o vídeo institucional de fim de ano do Centro de Tecnologia da UFSM<sup>1</sup>.

Neste primeiro vídeo pode-se notar pequenas falhas durante a execução dos movimentos, mostrando a necessidade de refinamento do algoritmo, entretanto toda a fala é contemplada com movimentos e olhos adequados às emoções selecionadas pelo algoritmo.

Em seguida, após refinamento do algoritmo, foram feitos cinco vídeos promocionais de fim de ano para a empresa Qiron Robotics<sup>2</sup>. Nessa sequência de vídeos os movimentos são mais fluidos, não havendo muitas pausas entre um movimento e outro. As representações de olhos também são adequadas ao conteúdo de fala do robô, o que transmite ainda mais a sensação de vida e credibilidade ao espectador.

Para melhor avaliação estética foram coletados os dados obtidos no questionário de avaliação das emoções. Através das taxas de acerto do questionário podemos avaliar se o robô consegue, ou não, transmitir corretamente as emoções. Um resumo dos resultados obtidos pode ser observado na matriz de confusão apresentada na Figura 5.1.

---

<sup>1</sup>Disponível em: [https://youtu.be/yfPBDu9\\_\\_lk](https://youtu.be/yfPBDu9__lk)

<sup>2</sup>Disponível em: [https://youtu.be/qVhA8I07R\\_I](https://youtu.be/qVhA8I07R_I)

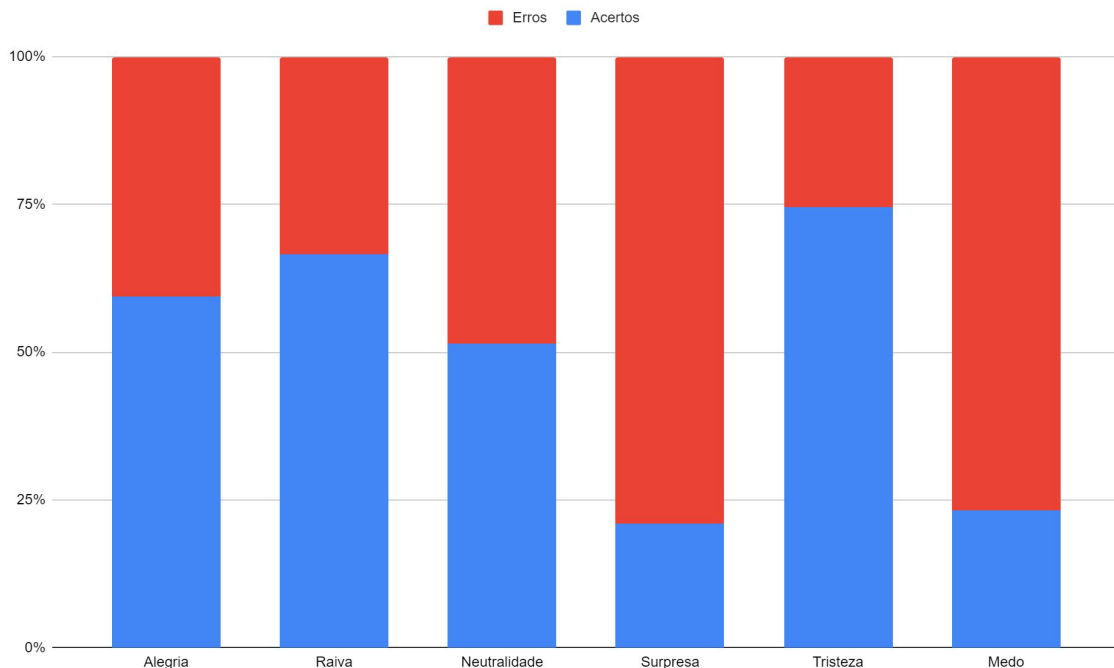
Figura 5.1 – Matriz de Confusão

	Alegria	Raiva	Neutralidade	Surpresa	Tristeza	Medo	Nenhuma
Alegria	<b>82 (59,4%)</b>	40 (29%)	4 (2,9%)	4 (2,9%)	0 (0%)	0 (0%)	8 (5,8%)
Raiva	2 (1,4%)	<b>92 (66,7%)</b>	8 (5,8%)	10 (7,2%)	7 (5,1%)	5 (3,6%)	14 (10,1%)
Neutralidade	15 (10,9%)	1 (0,7%)	<b>71 (51,4%)</b>	8 (5,8%)	4 (2,9%)	0 (0%)	39 (28,3%)
Surpresa	46 (33,3%)	20 (14,5%)	5 (3,6%)	<b>29 (21,4%)</b>	2 (1,4%)	4 (2,9%)	32 (23,2%)
Tristeza	2 (1,4%)	4 (2,9%)	6 (4,3%)	5 (3,6%)	<b>103 (74,6%)</b>	11 (8%)	7 (5,1%)
Medo	10 (7,2%)	22 (15,9%)	3 (2,2%)	32 (23,2%)	2 (1,4%)	<b>32 (23,2%)</b>	37 (26,8%)
Nenhuma	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	<b>0 (0%)</b>

Fonte: Autora

Das seis emoções presentes no questionário, quatro delas foram reconhecidas com mais facilidade pelos participantes da pesquisa, são elas: neutralidade (51,4%), alegria (59,5%), raiva (66,7%) e tristeza (74,6%). Podemos ter uma visão geral das taxas de acerto e erro no reconhecimento das emoções na Figura 5.2.

Figura 5.2 – Taxa de reconhecimento das emoções



Fonte: Autora

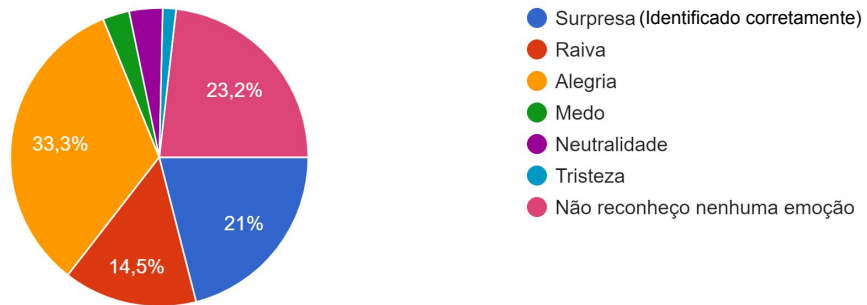
A surpresa (21%) e o medo (23,2%) se mostraram mais difíceis de serem reco-

nhecidas. Como podemos ver na Figura 5.3 a surpresa muitas vezes é confundida com alegria, ou até mesmo raiva, emoções de alta excitação.

Figura 5.3 – Taxa de reconhecimento da emoção surpresa

Surpresa

138 respostas



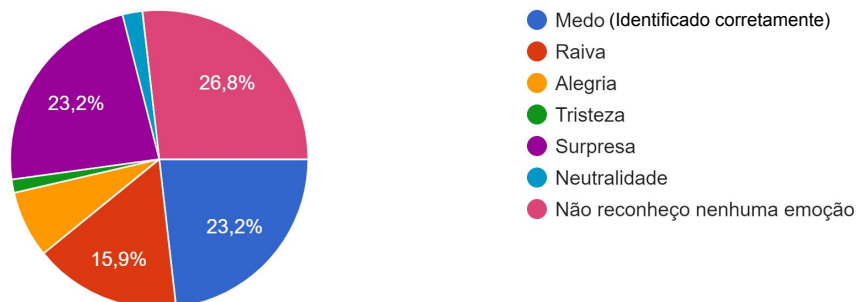
Fonte: Autora

Já o medo (Figura 5.4) muitas vezes foi confundido com surpresa e raiva. Houve também vários casos onde nenhuma emoção foi reconhecida.

Figura 5.4 – Taxa de reconhecimento da emoção medo

Medo

138 respostas

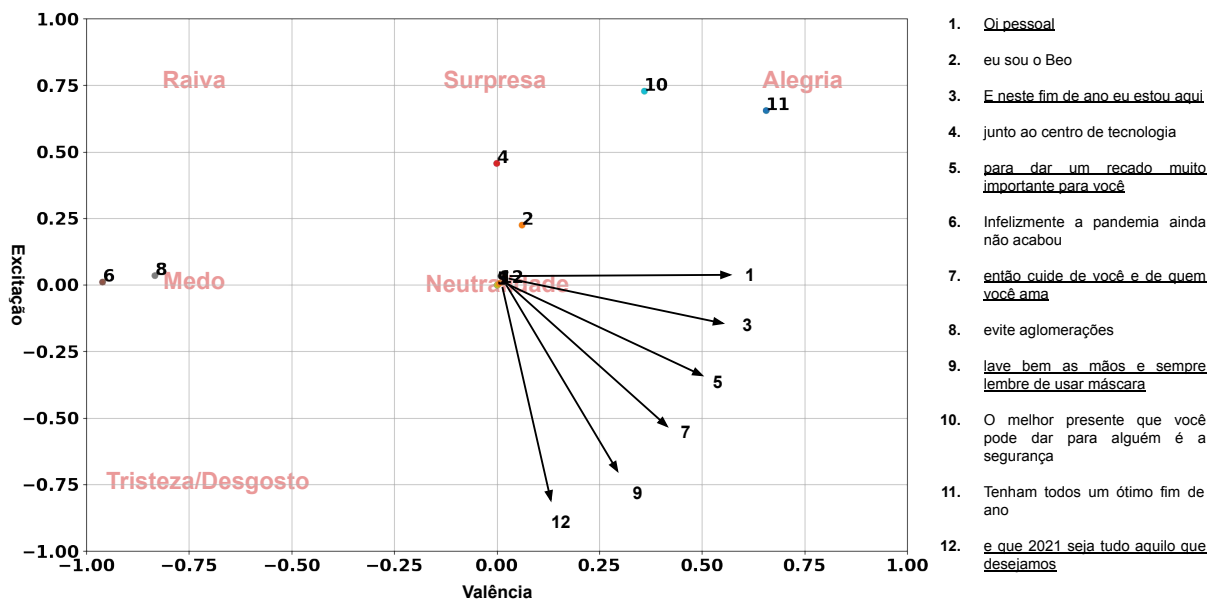


Fonte: Autora

## 5.2 MÉTODO PARA DETECÇÃO DE EMOÇÃO EM TEXTO

A partir dos cálculos da valência e da excitação das frases, foi possível estimar em quais regiões de emoção se encontra cada uma das frases que compõem um texto, utilizando a metodologia apresentada no capítulo 4.1 deste trabalho. Na Figura 5.5 podemos observar os resultados da classificação das frases do vídeo institucional de fim de ano do Centro de Tecnologia da Universidade Federal de Santa Maria.

Figura 5.5 – Classificação de frases transpostas para um plano 2-D

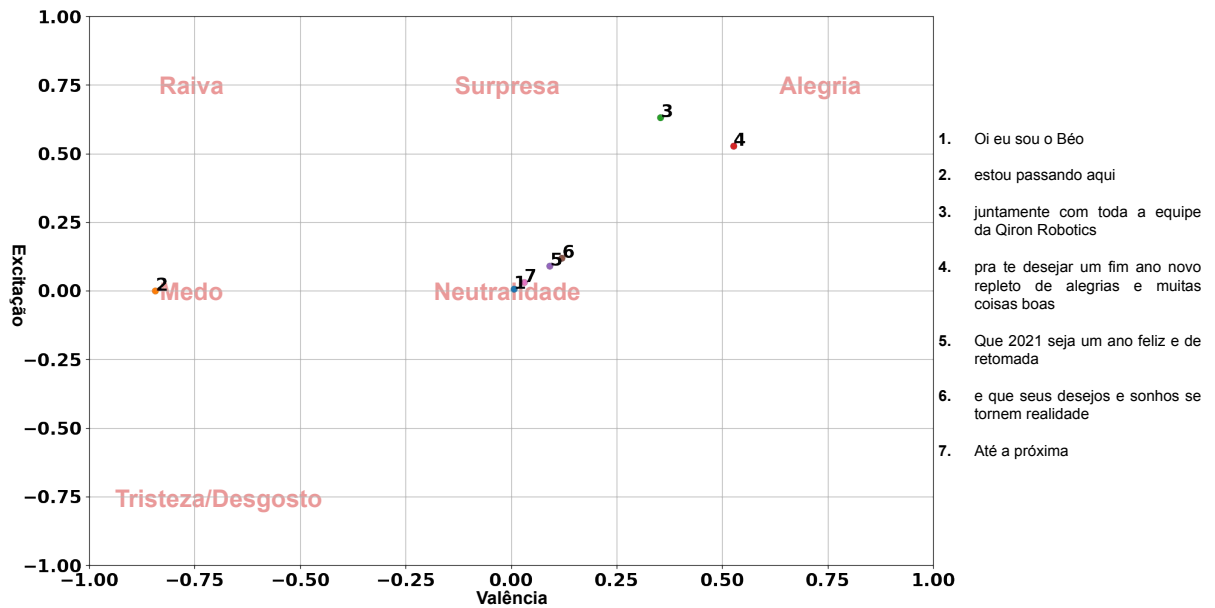


Fonte: Autora

Em rosa estão representadas as zonas de cada emoção. Notamos que metade das frases, sublinhadas na imagem, ficaram bem na região central do gráfico, com relação “x” e “y” muito próximas a zero, ou seja na região de neutralidade. Vale ressaltar que o vídeo institucional de fim de ano do Centro de Tecnologia foi totalmente produzido de forma automática, utilizando apenas o algoritmo gerador de apresentações desenvolvido durante este projeto.

Em seguida, serão apresentadas algumas classificações de textos modelos utilizados na empresa Qiron Robotics, como a mensagem de fim de ano da empresa. É possível observar (Figura: 5.6) que a maior parte das frases fica na região neutra com leves tendências para a alegria, o que condiz com o conteúdo do texto em questão. Porém, também podemos notar uma inconformidade com a classificação da frase número 2, sendo esta categorizada na região de medo do gráfico.

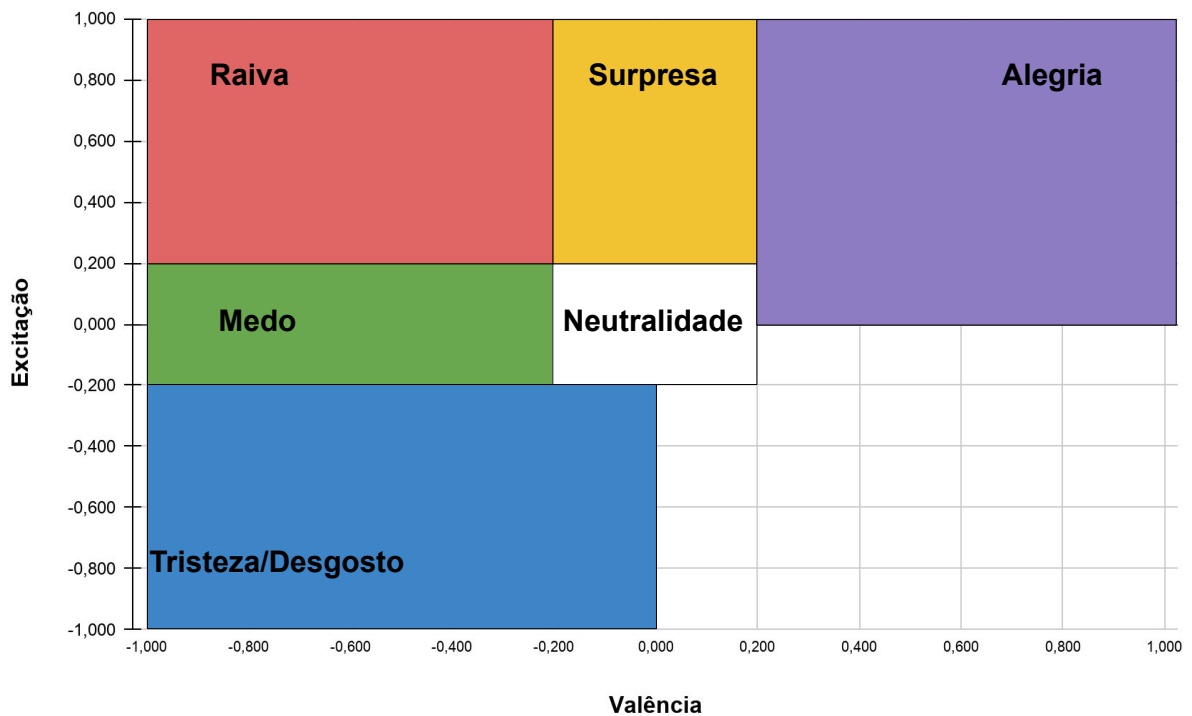
Figura 5.6 – Classificação de frases transpostas para um plano 2-D



Fonte: Autora

A classificação geral da frase, para escolha do gesto que irá representar a emoção, é feita de acordo com a região do gráfico na qual a frase ficou localizada, como mostrado na Figura 5.7

Figura 5.7 – Classificação de frases por região do gráfico



Fonte: Autora

Essas classificações foram usadas para a geração dos movimentos dos vídeos apresentados, atualmente o algoritmo de detecção de emoções no texto passa por um processo de refinamento, então prevemos uma significativa melhora em seu desempenho.

### 5.3 LOCAL DE APLICAÇÃO

Além de vídeos promocionais e institucionais, o algoritmo pode ser usado para criação de diversas outras apresentações. Dessa forma a Qiron Robotics promoveu o desenvolvimento de um ambiente gráfico (Figura 5.8) onde os clientes da empresa podem utilizar o algoritmo sem a necessidade de conhecimento prévio de linguagem de programação.

Figura 5.8 – Aplicação gráfica

OPERAÇÕES	NOME	IDIOMA	TEMPO (S)
🔍 🗑️	institucional_02_s1	pt-br	0:00:02.088000
🔍 🗑️	institucional_02_s10	pt-br	0:00:04.272000
🔍 🗑️	institucional_02_s11	pt-br	0:00:06.192000
🔍 🗑️	institucional_02_s12	pt-br	0:00:06.024000
🔍 🗑️	institucional_02_s13	pt-br	0:00:03.072000
🔍 🗑️	institucional_02_s2	pt-br	0:00:04.320000
🔍 🗑️	institucional_02_s3	pt-br	0:00:04.104000
🔍 🗑️	institucional_02_s4	pt-br	0:00:04.824000
🔍 🗑️	institucional_02_s5	pt-br	0:00:06.676000
🔍 🗑️	institucional_02_s6	pt-br	0:00:04.848000

Fonte: Autora

Na figura acima o sistema conta com um campo para seleção da emoção que o movimento deve representar, isso devido a algumas limitações do servidor para tratar com os códigos relativos à parte de detecção automática da emoção no texto. Nesse ambiente o movimento, então, é criado a partir de uma fala previamente adicionada e da seleção da emoção desejada pelo usuário. O movimento e os olhos gerados (Figura 5.9), juntamente com a fala, formam a apresentação final.

Figura 5.9 – Composição da Apresentação



Fonte: Autora

Na primeira linha de tempo apresentada anteriormente são alocadas as células de movimentos, na segunda, as falas, enquanto na terceira, as células das imagens que irão representar nos olhos. Dessa forma é possível que o usuário faça alterações das células inseridas, arrastando-as de um lado ao outro no correr da linha de tempo da apresentação.

## 6 CONCLUSÃO

No caso deste projeto, o principal objetivo era promover um sistema que fosse capaz de detectar as emoções em texto gerar apresentações completas, de textos, que fossem capazes de promover empatia, validando os conceitos relacionados à demonstração de emoções por parte de robôs humanoides. Neste capítulo serão apresentadas as avaliações dos resultados obtidos durante a execução, bem como possíveis problemas de execução e suas possíveis soluções.

Os vídeos usados para a validação deste trabalho foram postados em redes sociais e apresentaram boa aceitação por parte do público. Em todas as apresentações o robô representou cenas fluídas, que não destoavam com suas falas.

Ainda assim há a necessidade de melhora nos algoritmos de geração de cenas, não apenas na melhora do seu desempenho, tornando-o um algoritmo mais compacto e que exija menos processamento, mas também fazendo melhoras no seu funcionamento, como por exemplo adicionar uma forma de gerar os gestos que compõem a cena. Atualmente os gestos advêm de um banco de dados gerados anteriormente. O banco de dados de gestos conta com um total de 66 gestos. Como o algoritmo foi pensado de forma a gerar cenas independentes para cada frase do texto, não há limitação do número de caracteres do texto a ser representado.

Com a aplicação do questionário ficou claro que algumas emoções são facilmente reconhecidas, entretanto outras, como medo e surpresa mostram resultados mais dispersos. Alguns *feedbacks* também vão ao encontro de que outras emoções podem ser adicionadas aos estudos. Algumas pessoas relataram ter reconhecido como "descontentamento", ou outra emoção negativa, a representação da "tristeza". Como não haviam outras opções no questionário a opção "não reconheço nenhuma emoção" foi marcada. Isso nos mostra que existe uma gama de outras emoções que podem ser adicionadas ao estudo, essas variações podem ser alcançadas através da variação da valência e da excitação dos movimentos. Tanto os olhos quanto os movimento corporais foram fundamentais para o reconhecimento das emoções.

O método de detecção de emoção em texto está passando por um processo de refinamento e, assim que terminado, será aplicado nas utilizações deste trabalho pela empresa Qiron Robotics. As melhoras no método não apenas estão sendo feitas para tornar o processo menos oneroso, mas também em vias de melhorar os resultados. Algumas frases, que tinham conotação positiva tinham classificação neutra ou até mesmo de valência negativa, o que acabava gerando cenas que representavam tristeza ou medo. No geral, quando inseridas em uma grande apresentação essas cenas não destoavam do geral, devido a mescla de gestos, porém é algo a ser corrigido.

De modo geral as respostas obtidas foram positivas, mostrando que o banco de



dados é efetivo em demonstrar a maioria das emoções, entretanto, a dificuldade no reconhecimento da surpresa e do medo, bem como o constatação de outras emoções, que não as estudadas por Ekman, comprovam que esta é uma área de estudo extremamente ampla e complexa. Dessa forma verificamos a necessidade da continuidade deste trabalho para a conquista de resultados superiores, tanto para a obtenção de melhores formas de representação das emoções em robôs, quanto para a melhora no algoritmo de detecção de emoções em texto.

## REFERÊNCIAS BIBLIOGRÁFICAS

ALEXANDERSON, S. et al. Style-controllable speech-driven gesture synthesis using normalising flows. **Computer Graphics Forum**, v. 39, n. 2, p. 487–496, 2020.

ASUS. **Tinker Board**: Asus tinker board - um computador de placa única baseado em arm. ASUS, 2020. Acesso em 06 de novembro de 2020. Disponível em: <<https://www.asus.com/us/Networking-IoT-Servers/AIoT-Industrial-Solutions/Tinker-Board-Series/Tinker-Board/>>.

BATES, J. **The Role of Emotion in Believable Agents**. Pittsburgh, PA 15213, 1994. 1-9 p. Acesso em 06 de novembro de 2020.

BECK, A. et al. Interpretation of emotional body language displayed by robots. **AFFINE '10: Proceedings of the 3rd international workshop on Affective interaction in natural environments**, p. 37–42, 2010.

BINALI, H.; WU, C.; POTDAR, V. Computational approaches for emotion detection in text. **4th IEEE International Conference on Digital Ecosystems and Technologies**, p. 172–177, 2010.

BREAZEAL, C. **Emotion and Sociable Humanoid Robots**. 77 Massachusetts Ave, NE18-5FL Cambridge, MA 02139, 2002. 1-42 p. Acesso em 28 de Outubro de 2020. Disponível em: <<https://web.media.mit.edu/~cynthiab/Papers/Breazeal-ijhcs02-final.pdf>>.

BRECAILO, S. de F. **Expressão Facial e Corporal na comunicação em LIBRAS**. [S.l.], 2012. 1-12 p. Acesso em 06 de novembro de 2020.

Carlos Germano. **Ensino de robótica aumenta capacidade de solucionar problemas, diz professora da UFPB**. Ascom/UFPB, 2020. Acesso em 27 de Maio 2021. Disponível em: <<https://www.ufpb.br/ufpb/contents/noticias/ensino-de-robotica-aumenta-capacidade-de-solucionar-problemas-diz-professora-da-ufpb>>.

CASTRO, B. et al. **BUTIA LARC/CBR 2019 RoboCup@Home Team Description Paper**. [S.l.], 2019. 7 p. Acesso em 28 de Outubro de 2020.

DU, Y. et al. Distributing expressional faces in 2-d emotional space. **CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval**, p. 395400, 2007.

EKMAN, P. **Facial Expression**. [S.l.], 1977. 97-116 p. Acesso em 06 de novembro de 2020.

GOLDBERG, Y. **Neural Network Methods for Natural Language Processing**. Bar Ilan University: Morgan & Claypool Publishers, 2017. 285 p.

GUPDA, N.; GILBERT, M.; FABBRIZIO, G. D. Emotion detection in email customer care. **Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text**, p. 10–16, 2012.

HEIDER, F.; SIMMEL, M. An experimental study of apparent behavior. **The American Journal of Psychology**, v. 57, n. 2, p. 243–259, 1944.

ITOH, K. et al. Mechanical design of emotion expression humanoid robot we-4rii. In: ZIELISKA, T.; ZIELISKI, C. (Ed.). **Romansy 16. CISM Courses and Lectures**. [S.l.]: Springer, Vienna, 2006. p. 255–262.

Jacqueline Kovacs. **Robot gets seal of approval**. Toronto Star, 2014. Acesso em 07 de dezembro de 2020. Disponível em: <[https://www.thestar.com/life/breakingthrough/2014/06/09/robot\\_gets\\_seal\\_of\\_approval.html](https://www.thestar.com/life/breakingthrough/2014/06/09/robot_gets_seal_of_approval.html)>.

James Burch. **In Japan, a Buddhist Funeral Service for Robot Dogs**. National Geographic, 2020. Acesso em 07 de dezembro de 2020. Disponível em: <<https://www.nationalgeographic.com/travel/destinations/asia/japan/in-japan--a-buddhist-funeral-service-for-robot-dogs/>>.

JEONG, J.; YANG, J.; BALTES, J. Robot magic show: Human-robot interaction. **The Knowledge Engineering Review**, v. 35, p. e15, 2020.

KOZIMA, H.; MICHALOWSKI, M. P.; NAKAGAWA, C. Keepon. **International Journal of Social Robotics**, v. 1, p. 3–18, 2009.

LARSEN, M. E. et al. We feel: mapping emotion on twitter. **IEEE journal of biomedical and health informatics**, v. 19, n. 4, p. 12461252, 2015.

MEDHAT, W.; HASSAN, A.; KORASHY, H. Sentiment analysis algorithms and applications: A survey. **Ain Shams Engineering Journal**, v. 5, n. 4, p. 1093–1113, 2014.

OLIVEIRA, M. G. de. **Um Estudo Sobre a Influência de Eventos Históricos Recentes nos Sentimento Expressados em Mensagens de Redes Sociais**. [S.l.], 2021.

PARO Robots. **PARO Robots**: Paro therapeutic robot. PARO Robots, 2014. Acesso em 07 de dezembro de 2020. Disponível em: <<http://www.parorobots.com/>>.

Paul Ekman. **Emblematic Slips**: The meaning of our movements and gestures. Paul Ekman Group, 2020. Acesso em 06 de novembro de 2020. Disponível em: <<https://www.paulekman.com/blog/emblematic-slips/>>.

\_\_\_\_\_. **Universal Emotions**: What are emotions? Paul Ekman Group, 2020. Acesso em 06 de novembro de 2020. Disponível em: <<https://www.paulekman.com/universal-emotions/>>.

POULIN, C. et al. Predicting the risk of suicide by analyzing the text of clinical notes. **PLoS ONE**, v. 9, n. 1, p. e85733, 2014.

Qiron Robotics. **Mais engajamento com o Beo**: Beo. Qiron Robotics, 2020. Acesso em 06 de novembro de 2020. Disponível em: <<https://alugueobeo.com.br/>>.

RIZZOLATTI, G.; ARBIB, M. A. Language within our grasp. **Trends in Neurosciences**, v. 21, n. 5, p. 188–194, 1998.

RIZZOLATTI, G.; CRAIGHERO, L. The mirror-neuron system. **Annual Review of Neuroscience**, v. 27, p. 169–192, 2004.

Robotis. **XL320**: Especifications. Robotis e-Manual, 2020. Acesso em 06 de novembro de 2020. Disponível em: <<https://manual.robotis.com/docs/en/dxl/x/xl320/>>.

RobotLAB. **Robô Humanoide NAO V6**: Robôs para desenvolvedores. RobotLAB, 2020. Acesso em 06 de novembro de 2020. Disponível em: <<https://www.robotlab.com/loja/robo-nao-programador>>.

RUSSEL, J. A.; MEHRABIAN, A. Evidence for a three-factor theory of emotions. **Journal of Research in Personality**, v. 11, p. 273–294, 1977.

SEYEDITABARI, A.; TABARI, N.; ZADROZNY, W. Emotion detection in text: a review. **CoRR**, abs/1806.00674, 2018.

SILVERA-TAWIL, D.; BROWN, S. A. Cross-collaborative approach to socially-assistive robotics: A case study of humanoid robots in a therapeutic intervention for autistic children. In: \_\_\_\_\_. **Social Robots: Technological, Societal and Ethical Aspects of Human-Robot Interaction**. Springer: Prentice Hall, 2019. cap. 9, p. 165–186.

SoftBank Robotics. **Pepper**. SoftBank Robotics, 2020. Acesso em 07 de dezembro de 2020. Disponível em: <<https://www.softbankrobotics.com/emea/en/pepper>>.

SONY. **aibo**: unleash wonder. SONY, 2020. Acesso em 07 de dezembro de 2020. Disponível em: <<https://us.aibo.com/>>.