

# Componentes com capacidade de Reconhecimento em Interação Humano Robô

Bolsista Gustavo Henrique de Oliveira (CTI) [gustavo.oliveira@cti.gov.br](mailto:gustavo.oliveira@cti.gov.br)

## Resumo

*Este trabalho incorpora atividades relacionadas aos Componentes com capacidade de Reconhecimento em Interação Humano Robô, reconhecimento de voz, reconhecimento de pessoas, reconhecimento de expressões faciais e corporais, entre outros, utilizando softwares disponíveis para tais finalidades e realizar integração dessas capacidades ao arcabouço do robô recepcionista do Núcleo de Robótica e Visão Computacional - NRVC/CTI. Estas atividades contribuem no desenvolvimento ao projeto “Informações não verbais aplicadas a um robô recepcionista”, projeto na área de Interação Humano Robô, que vem sendo desenvolvido pelo Núcleo.*

*Palavras-chave: Interação Humano Robô, imagens, reconhecimento.*

## 1. Introdução

O objetivo geral associado ao projeto proposto é contribuir para a pesquisa na área de Interação Humano Robô – IHR, do inglês HRI - Human-Robot Interaction é fundamental para o desenvolvimento de técnicas procurando aumentar a proximidade e interação entre robôs e humanos. Seu objetivo é permitir uma interação natural e efetiva entre humanos e robôs, definindo princípios e algoritmos através de conferências realizadas pela comunidade IHR - HRI Conference [HRI, 2019], através do projeto “Informações Não-Verbais na Interação Humano Robô Aplicado a um Robô Recepcionista” do Núcleo de Robótica e Visão Computacional – do Centro de Tecnologia da Informação Renato Archer – NRVC/CTI, associado à área de IHR tem como objetivo abordar tecnologias, promover o conhecimento e disseminação da problemática associada a essa área.

O framework [PINHEIRO, 2018] desenvolvido pelo NRVC/CTI, denominada como cérebro, é baseado no sistema herdado de um projeto realizado pela Carnegie Mellon University – CMU, com orientação do Prof. Reid Simmons no grupo de robótica, com a finalidade desenvolver um robô recepcionista.

A Figura 1 apresenta os principais subsistemas associados ao framework [PINHEIRO, 2016] desenvolvido no CTI do robô recepcionista com capacidade de interação não verbal. Nessa figura destacam-se os seguintes subsistemas:

- Detecção e reconhecimento de voz: sua função é transformar a voz em texto.
- Entrada por texto: Utiliza um teclado posicionado próximo ao robô, o usuário pode fazer requisições em forma de texto.
- Analisador de texto: Sua função é quebrar a sentença de entrada do usuário, criando um mapa léxico das palavras faladas ou digitadas, que é utilizado pela máquina de regras verbais.
- Máquina de regras verbais: Sua Função é gerar as respostas em função do contexto das frases do usuário, procurando informações na base de conhecimento local, ou externa na DBPedia e Google News.
- Gerenciamento do Conhecimento: trata-se da implantação de um banco de conhecimento orientado a ontologia composto por informações próprias do domínio

de um recepcionista e por conceitos mais gerais.

- Comunicação não verbal: implementa a comunicação não verbal a partir de informações visuais, com a presença do usuário, reconhecimento facial e reconhecimento de emoções.
- Timeline: É utilizado o conceito de Timeline ou linha do tempo, como a estrutura de dados capaz de armazenar as interações realizadas. Cada interação humano-robô cria um slot que é anexado à linha do tempo e que concentra dados sobre o usuário da interação, suas perguntas verbais e emoções registradas.
- Máquina de regras não verbais: Contém as regras de ação vinculadas às informações não verbais, como as regras disparadas com o aparecimento do usuário, as regras disparadas com a saída do usuário, regras disparadas durante a interação com o usuário e regras de movimentação randômica quando nenhum usuário está presente.
- Máquina de respostas: Recebe as respostas das máquinas de regras (verbal e não verbal), formata a resposta e repassa para os dispositivos de saída.
- Dispositivos de saída: corresponde atualmente ao dispositivo que realiza a interação com o usuário, sintetizando voz e realizando movimentos correspondentes a fala. Pode ser um Avatar ou robô humanoide.
- Avatar, o avatar do robô inicialmente foi utilizado o mesmo do sistema legado da CMU, e atualmente foi substituído por uma avatar criado no NRVC/CTI. Onde foi desenvolvido um servidor de mensagens, tornando o avatar capaz de receber mensagens vindas da Máquina de Resposta, executando as ações desejadas, como: expressão a ser realizada pelo avatar, texto a ser falado e posição de seu rosto. Um protocolo de mensagem, para representar essas ações, foi criado utilizando o formato JSON [JSON,2019], e a escalabilidade do sistema permite que novos avatares ou robôs sejam facilmente integrados no sistema atual.

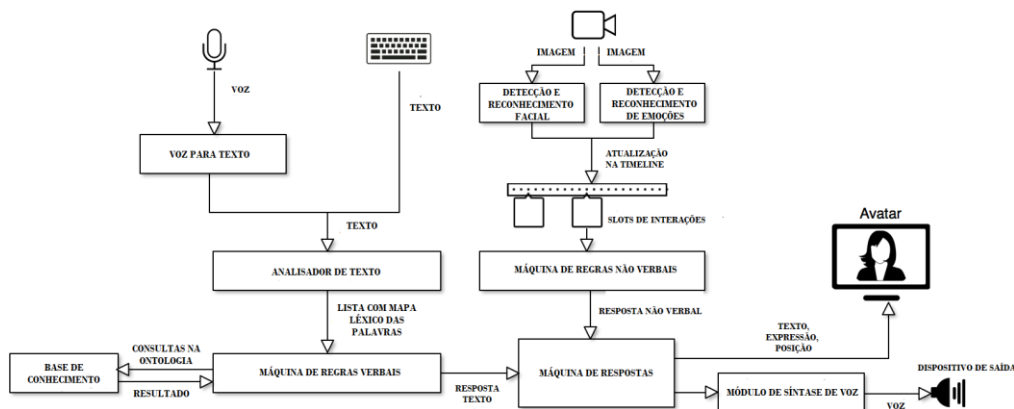


Figura 1: Arquitetura do Framework do Robô Recepcionista

O sistema foi desenvolvido utilizando J2EE, Framework Spring MVC, JSP e servidor de aplicação Tomcat, permitindo ser utilizado por terminais remotos.

Este artigo apresenta às atividades realizadas, na Seção 2.1 a integração arcabouço do robô recepcionista ao avatar desenvolvido no CTI, a Seção 2.2. a integração do robô NAO ao arcabouço do robô recepcionista, na Seção 2.3 demonstra a evolução da integração do sistema de reconhecimento facial fornecido pela Neurotechnology ao arcabouço do robô recepcionista, a Seção 2.4 apresenta experimentos para automatização de ligações para ramais e finalmente a Seção 3 apresenta as conclusões.

## 2. Atividades Realizadas

### 2.1 Integração do arcabouço do robô recepcionista ao Avatar

Inicialmente foi integrado o avatar herdado do projeto da CMU ao sistema desenvolvido no NRVC/CTI, representado na Figura 3. Sendo que mais tarde criou-se um novo avatar, com características de recepcionistas brasileiras. Para o projeto foi criado um avatar baseado na plataforma de desenvolvimento de jogos multiplataforma Unity [UNITY, 2019], apresentado na Figura 4.

Para a integração entre o avatar unity com o cérebro são estabelecidas conexões baseadas no protocolo TCP/IP. Com a conexão estabelecida o cérebro envia instruções utilizando estruturas JSON, que constituída em uma coleção de pares nome/valor, representada nos seguintes formatos:

a) {sentence, iduser}, sendo que:

- sentence é uma string que será sintetizada em voz para ser falada pelo avatar;
- iduser é um número inteiro que representará a identificação do usuário atualmente se comunicando com a recepcionista.

A partir do momento que o avatar recebe este JSON, ele faz o processamento da informação e externaliza na forma de fala. Quando o avatar inicia o processo de fala, ele envia ao cérebro um comando "START TALKING", com isso o robô recepcionista fica no estado de surdez, para que não ouça a própria voz e processe uma nova informação. Ao finalizar a fala o avatar envia ao cérebro o comando "FINISH TALKING", onde o cérebro volta ao seu estado inicial, podendo receber novos comandos de fala do usuário.

b) {expression,starx,stary,starz}, onde:

- expression é uma string que indicará a expressão que o avatar deverá fazer;
- starx/stary/starz são números do tipo float que representará o ponto no espaço tridimensional do Unity que o avatar deverá focar.

Este JSON é responsável pelos movimentos e expressões do avatar. Através destes comandos o avatar consegue seguir o usuário de acordo com a posição que ele está em relação à câmera. O comando para o avatar seguir o usuário é enviado em tempo real. As informações são enviadas entre o cérebro e o avatar em conexões distintas com o intuito do avatar seguir o usuário e falar simultaneamente, sem que um comando atrapalhe a execução do outro.

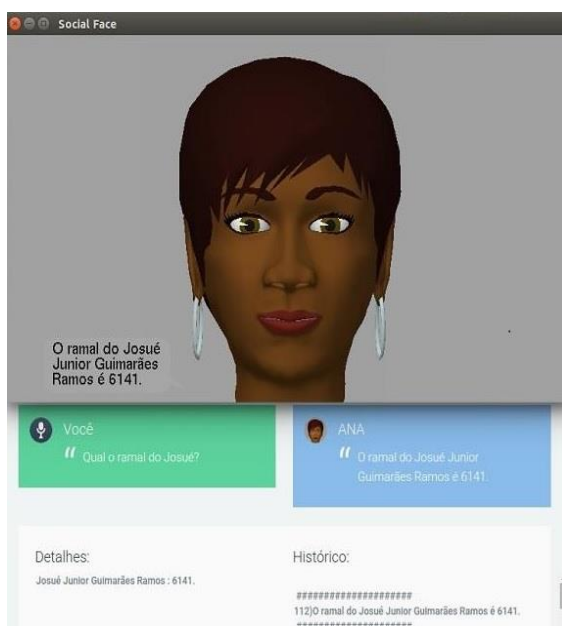


Figura 2 - Interface desenvolvida no CTI e Avatar herdado da CMU



Figura 3 - Interface e Avatar desenvolvidas no CTI

## 2.2 Integração do arcabouço do robô recepcionista ao robô NAO

A Integração entre o robô NAO e o arcabouço do robô recepcionista segue um processo evolutivo com complexidade crescente e utiliza o protocolo TCP/IP para conectar diferentes processos, seguindo a mesma forma que o Avatar. A Figura 4 mostra esse esquema.

A primeira solução adotada para a prova de conceito para conexão entre o robô NAO e o sistema recepcionista consistiu na substituição direta do Avatar pelo robô, sendo que parte do sensoriamento é feito pelo robô para a detecção de pessoas e acompanhamento destas e os demais componentes como microfone e câmera Intel RealSense 3D F200 [INTEL REALSENSE], conectados aos sistema do recepcionista. Isto permitiu a realização de interação básica substituindo o Avatar pelo robô NAO.

Uma segunda solução adotada substituí os dados recebidos da câmera Intel RealSense, por dados enviados pela câmera do próprio robô. A Figura 5 mostra o diagrama em Choreography [CHOREGRAPHE,2019], que é um aplicativo de desktop multiplataforma do NAO, que permite a criação de animações, comportamentos e diálogos; Com o aplicativo é possível realizar testes com o robô em ambiente simulado ou no robô real, monitorar e controlar o robô. Nesta Figura observa-se o comportamento “Face Tracker” que segue a face da pessoa presente, e aparece na região à direita a pose do robô e a pessoa que está sendo seguida, com a detecção da pessoa inicia a interação.

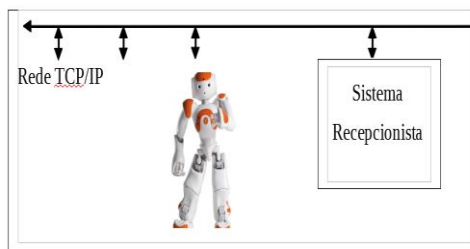


Figura 4 - Esquema de conexão entre o robô recepcionista e o NAO

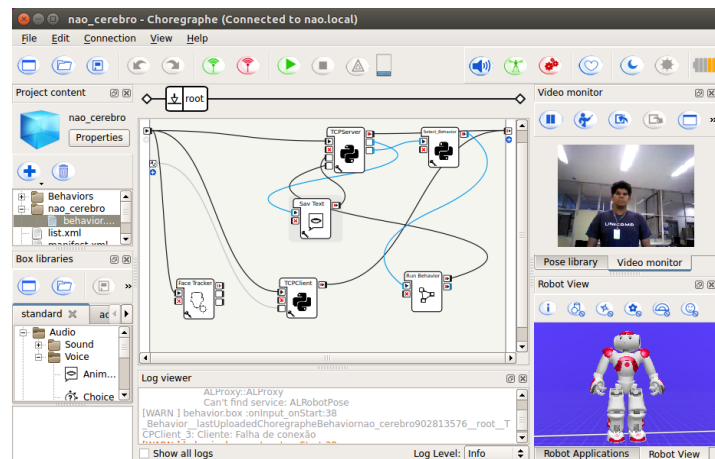


Figura 5 - Choregraphe com a interface do robô recepcionista

A segunda solução permite substituir o Avatar e a câmera Intel RealSense pelo robô e seus sensoriamentos, sendo que a fala e a detecção de pessoas e segui-las feito pelo robô, a captura de voz ainda continua sendo feita pelo sistema do recepcionista.

## 2.3 Integração SDK Neurotechnology ao arcabouço do robô recepcionista

Para o desenvolvimento e uso do sistema de reconhecimento facial da Neurotechnology [NEUROTECHNOLOGY, 2019] a empresa oferece durante um período de trinta dias uma versão “Trial”, após esse período é necessário à obtenção de licenças de uso, licenças que são validadas individualmente por computador.

Das linguagens de programação disponíveis na SDK, foi escolhida a linguagem Java para o desenvolvimento de um sistema de reconhecimento facial compatível com as necessidades encontradas no projeto, como base os exemplos disponíveis no kit de desenvolvimento.

A versão desenvolvida foi integrada ao cérebro do robô recepcionista do CTI com a utilização do protocolo de comunicação TCP/IP, escolhido para ser usado na comunicação entre o reconhecedor e o cérebro para evitar a perda de dados que transitam entre os dois sistemas, foi criado um cliente (Sistema de Reconhecimento Facial) e um servidor (Cérebro da recepcionista), onde o cliente envia informações obtidas no processamento da imagem. A principal informação que é enviada é o ID do usuário com maior score que foi localizado na base de dados.

Quando há a detecção do usuário, mas a imagem capturada não é de boa qualidade, o reconhecedor facial não pode gerar o template da imagem, devido as possíveis causas: iluminação do ambiente; ângulo da câmera; distancia do rosto em relação à câmera; ângulo do rosto em relação à câmera; resolução da imagem;

Com erro na geração do template não é possível fazer o processamento da imagem e o sistema reconhecimento facial envia uma requisição para a câmera capturar uma nova imagem e novamente ser feito processamento da mesma.

Após obtenção de uma imagem com a qualidade e requisitos necessários para o processamento, é gerado o template da imagem, que é um arquivo que contém dados referentes aos pontos característicos da face. Este template é utilizado para a busca na base dos usuários cadastrados, e retorna o ID e o score dos usuários com maior semelhança ao usuário que está sendo identificado. O cliente envia uma String contendo o ID do usuário, o ID do nome usuário conforme o cadastro feito e o score é um número que corresponde a semelhança dos usuários cadastrados em relação ao usuário que está sendo identificado.

Caso a pessoa não seja localizada na base de usuários cadastrados, o sistema automaticamente inicia o modo de cadastramento, onde são armazenados nas bases de imagens e templates os respectivos arquivos referentes ao novo usuário cadastrado. Se o usuário não desejar realizar seu cadastro no momento do cadastramento, o sistema gera um ID automático, sendo possível a alteração desse ID pelo nome do usuário em uma futura interação entre o novo usuário e o robô recepcionista.

A Figura 6 ilustra o processo de captura pela câmera, reconhecimento facial e comunicação com o cérebro do avatar. Na Figura 7 apresenta a interface da versão desenvolvida com base no SDK da Neurotechnology integrado com o arcabouço do robô recepcionista.

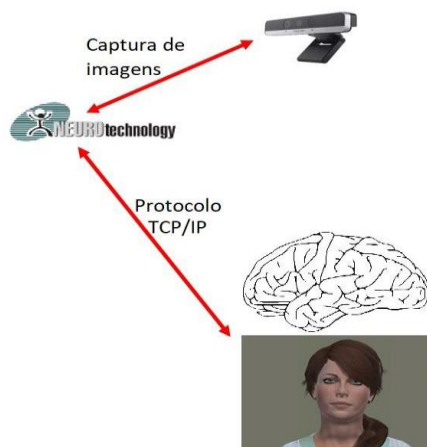


Figura 6: Processos para Reconhecimento Facial e comunicação com o Cérebro do Robô Recepcionista

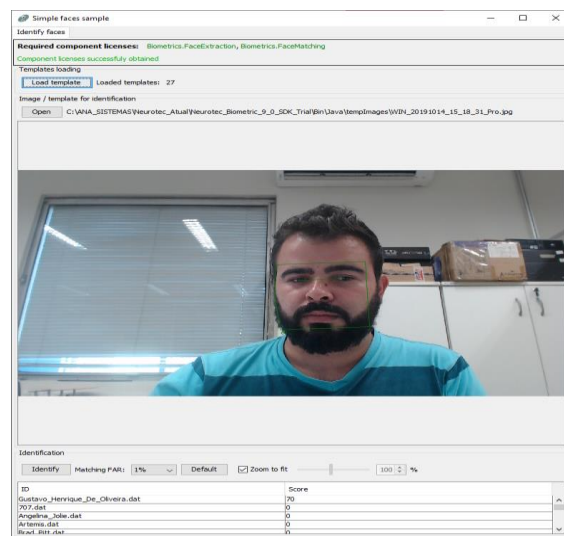


Figura 7: Interface da versão desenvolvida com SDK Neurotechnology

## 2.4 Experimentos com ligações automáticas para ramais

Uma das principais funções do robô recepcionista do CTI é informar ramais dos colaboradores do centro, com intuito de automatizar essa função o robô terá a capacidade de realizar a ligação para o ramal desejado e transferir para um telefone ao lado do robô, para que o usuário possa entrar em contato com a pessoa requerida. Para isso foram realizados diversos experimentos com o uso SDKs e alguns softphones de código aberto, como: Peers SDK [PEERS, 2019], Ozeki SDK [OZEKI, 2019], sipchat, linphone, voicechat, zoiper, X\_Lite.



Com o uso do Ozeki SDK, utilizando a função disponível para transferência de ligação, foi possível gerar uma versão, configurada de acordo com o sistema Voip utilizado no CTI, que através do protocolo de comunicação TCP/IP recebe o ramal do colaborador requerido e executa a ligação.

### 3. Conclusões

Apresentou-se as atividades desenvolvidas no programa PCI, cuja ênfase é no uso de componentes com capacidade de Reconhecimento em Interação Humano Robô.

Assim utilizou-se sistemas para reconhecimento de voz, faces, incluindo emoções e expressões faciais. A ênfase da maior parte das atividades esteve relacionada ao Avatar, o reconhecedor de faces da Neurotechnology e a integração dos mesmos ao arcabouço do robô recepcionista. Para escolha de tal, realizou-se um levantamento de sistemas comerciais que atendiam as necessidades de precisão e usabilidade requerida pelo cérebro do robô recepcionista e pelo projeto como um todo. A partir deste realizou-se diversos testes que verificavam a compatibilidade desses sistemas com a parte já implementada no robô.

Foi apresentada uma possível funcionalidade de uso do robô NAO, na atuação como recepcionista, que integrado com o framework do recepcionista do CTI pode suprir essa função.

Como objetivo da continuidade deste projeto, tem-se a realização de melhorias na interface gráfica de diálogo do robô com o usuário; integrações de novas funcionalidades, como ligações automáticas para ramais; melhorias no diálogo do robô; experimentos com o uso do robô recepcionista.

### Referências

- CHROREGRAPHE**, Choregraphe User Guide. Disponível em: <<http://doc.aldebaran.com/1-14/software/choregraphe/index.html>>. Acesso: 08 de outubro 2019.
- HRI**, Conference HRI. Disponível em: <<http://humanrobotinteraction.org/category/conference/>>. Acesso: 08 de outubro 2019.
- INTEL REALSENSE**, Intel RealSense 3D (F200). Disponível em: <<http://software.intel.com/en-us/realsense/f200camera/>>. Acesso: 15 de novembro 2016.
- JSON**, JSON - JavaScript Object Notation - Notação de Objetos JavaScript. Disponível em: <<http://www.json.org/json-pt.html>>. Acesso: 10 de fevereiro 2016.
- NEUROTECHNOLOGY**, Neurotechnology. Disponível em: <<http://www.neurotechnology.com/>>. Acesso em: 15 de maio 2016.
- OZEKI**, Ozeki Voip Sip SDK. Disponível em: <[http://www.voip-sip-sdk.com/p\\_290-ozeki-voip-sdk-voip.html/](http://www.voip-sip-sdk.com/p_290-ozeki-voip-sdk-voip.html/)>. Acesso em: 05 de março 2019.
- PEERS**, Peers Java SIP Softphone. Disponível em: <<http://peers.sourceforge.net/>>. Acesso em: 10 de fevereiro 2019.
- PINHEIRO, P. G.; RAMOS, J. J. G.; DONIZETE, V. L.; PICANCO, P.; OLIVEIRA, G. H.** Workplace Emotion Monitoring - An Emotion-Oriented System Hidden Behind a Receptionist Robot. In: Dan Zhang e Bin Wei. (Org.). Workplace Emotion Monitoring - An Emotion-Oriented System Hidden Behind a Receptionist Robot. 1ed.: Springer International Publishing, 2016, v. 1, p. 320-340.
- PINHEIRO, P. G.; RAMOS, J.J.G.; OLIVEIRA, G. H.** Smile and Talk - ANA, The Architecture of a Robot That Verbally and Nonverbally Understands You. In: Anais do XXII Congresso Brasileiro de Automática, 2018.
- UNITY**, Unity Manual. Disponível em: <<https://docs.unity3d.com/Manual/UnityManual.html>>. Acesso: 08 de outubro 2019.