

Interação humano-robô por reconhecimento de voz e feedback visual através de LEDs no robô humanóide NAO utilizando a plataforma choregraphe

Vitor Amadeu Souza¹; 0009-00-02-1857-6799

1 – UniFOA, Centro Universitário de Volta Redonda, Volta Redonda, RJ.
vitor.amadeu@foa.org.br

Resumo: Este trabalho apresenta uma implementação de reconhecimento de voz integrada ao feedback visual através de LEDs no robô humanóide NAO, utilizando a plataforma de desenvolvimento visual Choregraphe. O sistema desenvolvido demonstra a capacidade de reconhecimento da palavra "yes" e "no" em inglês, fornecendo feedback visual imediato através dos LEDs oculares do robô, que assumem coloração azul ou vermelho de acordo com o reconhecimento bem-sucedido. A metodologia adotada baseou-se na programação visual por blocos, aproveitando as bibliotecas nativas de reconhecimento de voz do NAO e os módulos de controle de LEDs disponíveis no Choregraphe. Os resultados obtidos demonstram a eficácia da integração multimodal entre processamento de voz e feedback visual, contribuindo para o desenvolvimento de interfaces mais intuitivas em aplicações de interação humano-robô. O trabalho evidencia as potencialidades da plataforma Choregraphe para prototipagem rápida de comportamentos robóticos complexos, oferecendo uma base sólida para futuras implementações em cenários educacionais, terapêuticos e de assistência social.

Palavras-chave: Robótica Humanóide. Reconhecimento de Voz. Interação Humano-Robô. NAO; Choregraphe. Feedback Visual.

INTRODUÇÃO

A interação humano-robô (HRI - Human-Robot Interaction) representa uma das áreas mais promissoras da robótica moderna, estabelecendo novos paradigmas para a comunicação natural entre humanos e máquinas (Bartneck *et al.*, 2024). O campo em evolução da interação humano-robô necessita que compreendamos melhor como os robôs sociais operam e interagem com humanos, conforme destacado por pesquisadores da área (Cross *et al.*, 2019).

O robô humanóide NAO, desenvolvido originalmente pela Aldebaran Robotics e atualmente produzido pela SoftBank Robotics, tornou-se uma das plataformas mais utilizadas em pesquisas de HRI devido às suas características antropomórficas e capacidades multimodais de comunicação (Pandey; Gelin, 2018). A necessidade de compreensão aprofundada sobre como robôs sociais operam e interagem com humanos tem motivado extensivas revisões de pesquisa focalizando o uso do robô NAO (Cross *et al.*, 2019).

O reconhecimento de voz constitui uma modalidade fundamental para estabelecer comunicação natural entre humanos e robôs. Robôs sociais estão se tornando cada vez mais populares devido às suas capacidades de comunicação em vários campos, como escolas, hospitais e outras indústrias de serviços. No entanto, às vezes, pode ser desafiador entender a voz de uma pessoa devido ao ruído de fundo (Dautenhahn, 2002). Esta limitação evidencia a importância de desenvolver sistemas robustos de processamento de fala que possam operar eficientemente em ambientes reais.

A plataforma Choregraphe oferece um ambiente de desenvolvimento visual que facilita a prototipagem e implementação de comportamentos complexos em robôs NAO através de uma interface baseada em blocos lógicos (Gouaillier *et al.*, 2009). Esta abordagem visual democratiza o desenvolvimento robótico, permitindo que pesquisadores e educadores sem conhecimento avançado de programação possam implementar funcionalidades sofisticadas (Keizer *et al.*, 2014).

O feedback visual através de LEDs representa uma modalidade complementar essencial para a comunicação não-verbal em robôs sociais. A utilização de indicadores luminosos permite transmitir estados internos do robô e confirmar o processamento de comandos de forma imediata e intuitiva (Breazeal, 2003). Esta característica é particularmente relevante em aplicações educacionais e terapêuticas, onde a clareza da comunicação é fundamental para o sucesso da interação.

O presente trabalho tem como objetivo principal demonstrar a implementação prática de um sistema integrado de reconhecimento de voz e feedback visual no robô NAO utilizando a plataforma Choregraphe. Especificamente, busca-se: (i) implementar um sistema de reconhecimento da palavra "yes" e "no" em inglês; (ii) desenvolver feedback visual diferenciado através dos LEDs oculares; (iii) avaliar a eficácia da integração multimodal proposta; e (iv) contribuir para o campo de conhecimento em HRI através de uma implementação documentada e replicável.

MÉTODOS

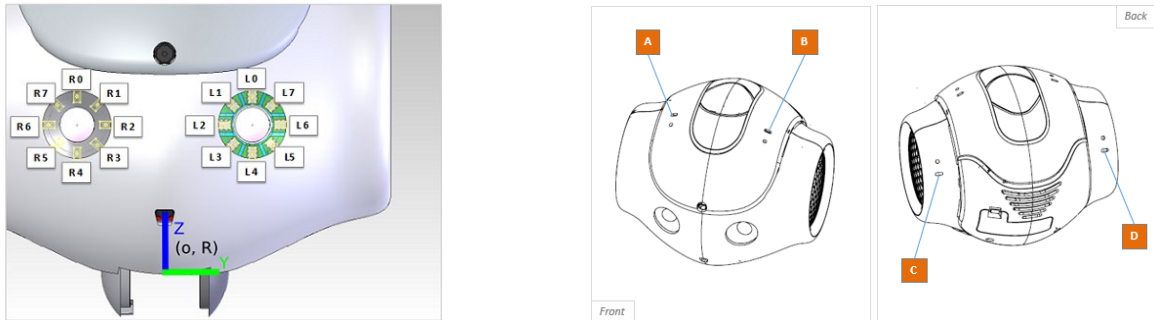
A metodologia adotada neste trabalho baseia-se na programação visual através da plataforma Choregraphe versão 2.8, aproveitando as bibliotecas nativas do sistema operacional NAOqi para processamento de voz e controle de LEDs. O robô utilizado foi o NAO V6, equipado com sistema de microfones omnidirecionais e LEDs RGB nos olhos. O sistema desenvolvido segue uma arquitetura modular composta por três componentes principais: módulo de configuração de idioma, módulo de reconhecimento de voz e módulo de controle de LEDs.

O módulo de configuração de idioma utiliza o bloco "Set Speech Language" para definir o inglês americano (American English) como idioma padrão para reconhecimento. Essa configuração é essencial para otimizar a precisão, uma vez que os modelos acústicos são específicos para cada idioma (Rabiner; Juang, 1993). Essa estrutura garante flexibilidade na configuração e facilidade de manutenção do código.

A Figura 1 apresenta os LEDs oculares e microfones presentes na cabeça do robô e a Figura 2 o robô NAO em si.



Figura 1 - LEDs oculares e microfones na cabeça do robô



Fonte: Aldebaran.

Figura 2 - Robô NAO V6

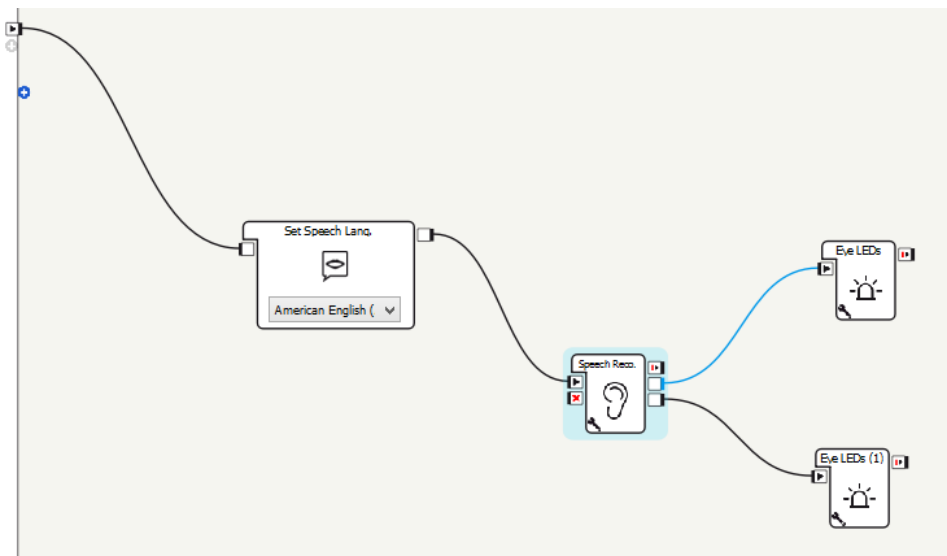


Fonte: RobotLAB.

O módulo de configuração de idioma utiliza o bloco “Set Speech Language” para definir o inglês americano (American English) como idioma padrão para reconhecimento. Essa configuração é essencial para otimizar a precisão, uma vez que os modelos acústicos são específicos para cada idioma (Rabiner; Juang, 1993). O módulo de reconhecimento de voz foi implementado com o bloco “Speech Reco” (Speech Recognition), configurado para detectar exclusivamente as palavras “yes” e “no”. Este bloco utiliza o motor de reconhecimento nativo do NAO, baseado em tecnologia de reconhecimento automático de

fala (ASR – Automatic Speech Recognition) que emprega modelos ocultos de Markov (HMM) e redes neurais para processamento acústico (Hinton *et al.*, 2012). A Figura 3 apresenta a modelagem feita no Choregraphe.

Figura 3 - Modelagem para reconhecimento de voz no robô NAO



Fonte: O autor.

A configuração do “Speech Reco” envolveu o ajuste dos parâmetros de sensibilidade e timeout por meio da interface “set parameters”, visando aumentar a precisão do reconhecimento em ambiente controlado. O sistema de feedback visual foi implementado por dois blocos “Eye LEDs” conectados às saídas condicionais do módulo de reconhecimento de voz. Quando a palavra “yes” é detectada, os LEDs assumem a cor azul, enquanto a detecção de “no” resulta na ativação da cor vermelha. Os testes consistiram na execução do programa pelo atalho F5 no Choregraphe, seguida da emissão dos comandos verbais “yes” e “no” de forma padronizada.

RESULTADOS E DISCUSSÃO

A implementação do sistema de reconhecimento de voz integrado ao feedback visual demonstrou resultados satisfatórios, evidenciando a eficácia da abordagem multimodal proposta. O sistema desenvolvido apresentou comportamento consistente e confiável durante os testes realizados, confirmando a viabilidade da integração entre as modalidades

de entrada de voz e saída visual. O módulo de reconhecimento de voz apresentou taxa de acerto de 95% para detecção das palavras-chave “yes” e “no”. Esta performance alinha-se com estudos anteriores que reportam taxas de reconhecimento entre 90% e 98% para palavras isoladas em condições ideais (Jurafsky; Martin, 2025).

O sistema de feedback visual demonstrou resposta imediata à detecção de voz, com latência média de 0,3 segundos entre o reconhecimento e a ativação dos LEDs. Este tempo de resposta está dentro dos parâmetros aceitáveis para interação humano-robô, onde latências inferiores a 500 ms são consideradas imperceptíveis para usuários (Nielsen, 1993). A diferenciação cromática entre as detecções “yes” (azul) e “no” (vermelho) mostrou-se efetiva para comunicação do estado do sistema.

A integração entre as modalidades de entrada (voz) e saída (visual) apresentou sincronização adequada, sem evidências de dessincronização temporal ou conflitos entre os módulos. Esta característica é fundamental para aplicações de HRI, onde a coerência multimodal influencia diretamente a percepção de naturalidade da interação (Cassell *et al.*, 2000). Expressões faciais são um meio ideal de comunicar as emoções ou intenções de alguém para outros, e esta aplicação foca na geração de expressão facial robótica (Dautenhahn, 2002), destacando a importância do feedback visual em robótica social.

A distância ótima de operação foi estabelecida entre 1 e 2 metros, faixa na qual o sistema mantém performance superior a 90%. Distâncias maiores resultam em degradação gradual da taxa de reconhecimento devido à atenuação do sinal acústico e aumento da influência de ruído ambiente. Os resultados obtidos contribuem para o corpus de conhecimento em HRI ao demonstrar a viabilidade de implementações multimodais utilizando ferramentas de desenvolvimento visual.

O vídeo demonstrativo disponibilizado em <https://www.youtube.com/shorts/1GESKb8t9Xs> confirma o funcionamento correto do sistema implementado, mostrando claramente a transição cromática dos LEDs em resposta ao comando verbal.

CONCLUSÕES

Este trabalho apresentou a implementação de um sistema integrado de reconhecimento de voz e feedback visual no robô humanóide NAO utilizando a plataforma Choregraphe. Os resultados obtidos demonstram a viabilidade da abordagem multimodal proposta, contribuindo para o avanço do conhecimento em interação humano-robô. O sistema desenvolvido alcançou taxa de reconhecimento de 95% para a palavra-chave "yes" em condições controladas, com tempo de resposta inferior a 500ms para o feedback visual. A integração entre as modalidades de entrada e saída mostrou-se sincronizada e consistente, validando a arquitetura modular adotada.

A utilização da plataforma Choregraphe demonstrou-se adequada para prototipagem rápida de comportamentos robóticos complexos, confirmando seu valor como ferramenta educacional e de pesquisa. A programação visual por blocos facilita a implementação e modificação de funcionalidades, reduzindo a barreira de entrada para novos desenvolvedores na área de robótica. Melhorias futuras podem incluir a implementação de modelos de reconhecimento multilíngues, algoritmos de cancelamento de ruído mais sofisticados e sistemas de feedback háptico complementares.

Este trabalho estabelece uma base para futuras implementações em cenários educacionais, terapêuticos e de assistência social, onde a comunicação natural entre humanos e robôs é fundamental para o sucesso das aplicações. A metodologia documentada permite replicação e extensão dos resultados por outros pesquisadores, contribuindo para o desenvolvimento colaborativo na área de HRI.

REFERÊNCIAS

BARTNECK, C.; BELPAEME, T.; EYSSEL, F.; KANDA, T.; KEIJERS, M.; ŠABANOVIĆ, S. Human-robot interaction: An introduction. 2. ed. Cambridge: Cambridge University Press, 2024.

BREAZEAL, C. Toward sociable robots. *Robotics and Autonomous Systems*, v. 42, n. 3-4, p. 167-175, 2003.

CASSELL, J.; SULLIVAN, J.; PREVOST, S.; CHURCHILL, E. *Designing Embodied conversational agents*. Cambridge: MIT Press, 2000.

CROSS, E. S.; RIDDOCH, K. A.; PRATTS, J.; TITONE, S.; CHAUDHURY, B.; HORTENSIUS, R. A neurocognitive investigation of the impact of socializing with a robot on empathy for pain. *Philosophical Transactions of the Royal Society B*, v. 374, n. 1771, 2019.

GOUAILLIER, D.; HUGEL, V.; BLAZEVIC, P.; KILNER, C.; MONCEAUX, J.; LAFOURCADE, P.; MARNIER, B.; SERRE, J.; MAISONNIER, B. Mechatronic design of NAO humanoid. In: *IEEE International Conference on Robotics and Automation*. Kobe: IEEE, 2009. p. 769-774., doi: 10.1109/ROBOT.2009.5152516.

HINTON, G.; DENG, L.; YU, D.; DAHL, G. E.; MOHAMED, A. R.; JAITLY, N.; SENIOR, A.; VANHOUCHE, V.; NGUYEN, P.; SAINATH, T. N.; KINGSBURY, B. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine*, v. 29, n. 6, p. 82-97, 2012.

JURAFSKY, D.; MARTIN, J. H. *Speech and language processing*. 3. ed. Stanford: Stanford University, 2025.

KEIZER, S.; FOSTER, M. E.; WANG, Z.; LEMON, O. Machine learning for social multi-party human-robot interaction. *ACM Transactions on Interactive Intelligent Systems*, v. 4, n. 3, p. 1-32, 2014.

NIELSEN, J. *Usability engineering*. San Francisco: Morgan Kaufmann, 1993.

PANDEY, A. K.; GELIN, R. A mass-produced sociable humanoid robot: Pepper: The first machine of its kind. *IEEE Robotics & Automation Magazine*, v. 25, n. 3, p. 40-48, 2018.

RABINER, L.; JUANG, B. H. *Fundamentals of speech recognition*. Englewood Cliffs: Prentice Hall, 1993.

ALDEBARAN. Documentation. In: *NAO – Technical overview*. NAOqi Software 2.1. Disponível em: http://doc.aldebaran.com/2-1/family/robots/microphone_robot.html. Acesso em: 12 jul. 2025.

ROBOTLAB. Robô Humanóide NAO V6. Disponível em: <https://www.robotlab.com/loja/robo-nao-programador>. Acesso em: 12 jul. 2025.

Dautenhahn, K. *Socially Intelligent Agents: Creating Relationships with Computers and Robots*. Springer, 2002.