

Implementação de síntese de voz em robôs humanoides: um estudo de caso com o robô NAO utilizando a biblioteca naoqi

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 artigo apresenta um estudo sobre a implementação de síntese de voz em robôs humanoides, utilizando como estudo de caso o robô NAO da Aldebaran Robotics. A pesquisa demonstra a aplicação prática da biblioteca NAOqi para programação de funcionalidades de texto-para-fala (Text-to-Speech - TTS), explorando os aspectos técnicos da interação humano-robô através da comunicação verbal. O trabalho analisa a implementação do módulo ALTextToSpeech, apresentando código Python para estabelecimento de conexão TCP/IP com o robô e execução de comandos de síntese de voz. Os resultados obtidos demonstram a eficácia da abordagem proposta para desenvolvimento de aplicações robóticas com capacidades de comunicação natural, contribuindo para o avanço da robótica social e sistemas de interação multimodal. A metodologia empregada baseia-se na análise experimental de código Python utilizando a classe ALProxy, permitindo controle remoto das funcionalidades vocais do robô NAO. As conclusões indicam que a biblioteca NAOqi oferece uma interface robusta e intuitiva para desenvolvimento de aplicações de síntese de voz, facilitando a criação de sistemas robóticos mais interativos e naturais na comunicação com usuários humanos.

Palavras-chave: Robótica humanoide. Síntese de voz. NAO. Interação humano-robô. NAOqi. Text-to-Speech.



INTRODUÇÃO

A robótica humanoide tem experimentado avanços significativos nas últimas décadas, particularmente no desenvolvimento de sistemas capazes de interagir naturalmente com seres humanos (Breazeal, 2003). A comunicação verbal representa um componente fundamental desta interação, permitindo que robôs transmitam informações, respondam a comandos e estabeleçam vínculos sociais com usuários (Fong; Nourbakhsh; Dautenhahn, 2003). Neste contexto, os sistemas de síntese de voz (Text-to-Speech - TTS) emergem como tecnologias essenciais para o desenvolvimento de robôs sociais eficazes.

O robô NAO, desenvolvido pela Aldebaran Robotics (posteriormente adquirida pela SoftBank Robotics), estabeleceu-se como uma plataforma amplamente utilizada em pesquisa e educação devido às suas capacidades avançadas de interação e programabilidade (Gouaillier *et al.*, 2009). Com aproximadamente 58 centímetros de altura e equipado com múltiplos sensores, atuadores e capacidades de processamento, o NAO oferece uma plataforma ideal para investigação de técnicas de interação humano-robô (Aldebaran Robotics, 2012), como mostra a Figura 1.

Figura 2 - Robô NAO V6



Fonte: RobotLAB.

A biblioteca NAOqi constitui o framework de desenvolvimento principal para programação do robô NAO, oferecendo uma arquitetura baseada em módulos que facilita o acesso às diversas funcionalidades do sistema (Santana; Santos; Barata, 2006). Entre os módulos disponíveis, o ALTextToSpeech destaca-se por permitir a implementação de capacidades de síntese de voz de alta qualidade, utilizando tecnologias avançadas de processamento de linguagem natural (Valagkouti *et al.*, 2022).

A importância da comunicação verbal em sistemas robóticos é amplamente reconhecida na literatura. Segundo Dautenhahn (2007), a capacidade de comunicação natural é fundamental para o desenvolvimento de robôs sociais que possam integrar-se efetivamente em ambientes humanos. Similarmente, Erol, Sevim-Cirak e Gülsoy (2023) argumentam que sistemas de TTS bem implementados podem significativamente melhorar a aceitação e eficácia de robôs em aplicações educacionais e terapêuticas.

O presente trabalho tem como objetivo principal analisar a implementação de funcionalidades de síntese de voz no robô NAO utilizando a biblioteca NAOqi, apresentando uma abordagem prática para desenvolvimento de aplicações robóticas com capacidades de comunicação verbal. Especificamente, pretende-se: (a) demonstrar o uso da classe ALProxy para estabelecimento de conexões com o robô NAO; (b) analisar a implementação do módulo ALTextToSpeech; (c) avaliar a eficácia da abordagem proposta através de testes práticos; e (d) discutir implicações para o desenvolvimento de sistemas robóticos interativos.

MÉTODOS

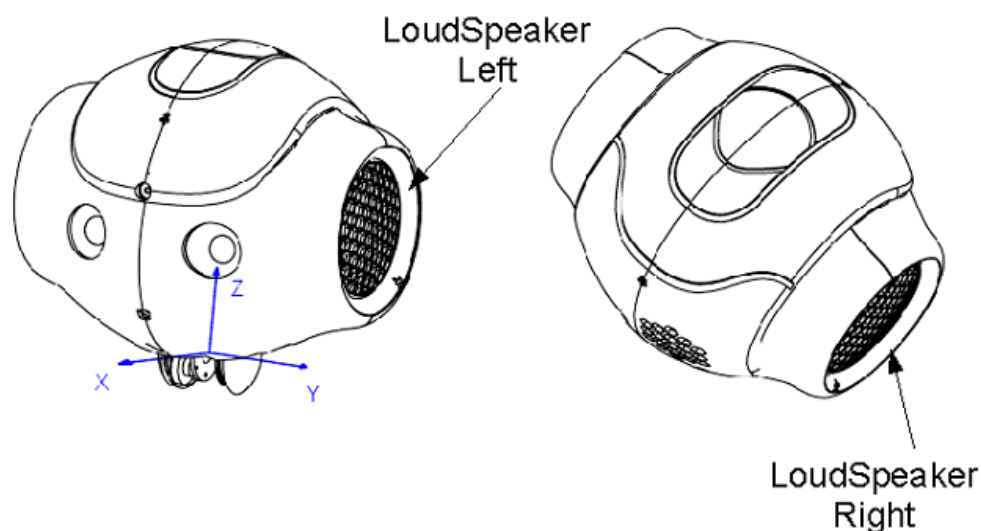
A metodologia empregada neste estudo baseia-se na análise experimental de implementações de síntese de voz utilizando o robô NAO e a biblioteca NAOqi. A abordagem adotada combina elementos de pesquisa aplicada e desenvolvimento experimental, seguindo princípios estabelecidos para investigação em robótica (Sierra-García *et al.*, 2023).

O ambiente experimental foi configurado utilizando um robô NAO V6 conectado a uma rede local via WiFi. O sistema de desenvolvimento consistiu em um computador executando Python 2.7 com a biblioteca NAOqi instalada. A comunicação entre o computador de

desenvolvimento e o robô foi estabelecida através do protocolo TCP/IP na porta padrão 9559, conforme especificações técnicas da plataforma (Aldebaran Robotics, 2014).

A implementação do sistema de síntese de voz foi realizada através da utilização da classe ALProxy da biblioteca NAOqi, especificamente acessando o módulo ALTextToSpeech. A arquitetura proposta baseia-se no padrão de proxy remoto, permitindo que aplicações Python executem comandos no robô NAO através de chamadas de rede (Santana; Santos; Barata, 2006). A Figura 2 mostra onde os alto-falantes do robô estão instalados.

Figura 2 - Speakers do Robô NAO



Fonte: ALDEBARAN ROBOTICS, 2012.

O código desenvolvido implementa uma função denominada `say_hello` que recebe como parâmetros o endereço IP do robô NAO (`nao_ip`) e a porta de comunicação (`nao_port`). Esta função cria uma instância da classe ALProxy especificando "ALTextToSpeech" como módulo alvo, estabelecendo assim a conexão com o subsistema de síntese de voz do robô.

A análise dos parâmetros de configuração focou nos aspectos críticos para estabelecimento de comunicação eficaz com o robô NAO. O parâmetro `nao_ip`, definido como "10.0.1.101" no exemplo estudado, representa o endereço IP do robô na rede local e deve ser adaptado conforme a configuração específica do ambiente de rede. A porta 9559 constitui a porta

padrão do serviço NAOqi e é utilizada para todas as comunicações com o broker de módulos do sistema operacional Gentoo embarcado no robô (Gouaillier *et al.*, 2009).

O protocolo de testes desenvolvido incluiu a execução sistemática do código de síntese de voz em diferentes condições operacionais. Os testes abrangeram verificação de conectividade de rede, latência de resposta do sistema, qualidade da síntese de voz e robustez da implementação. A frase "Hello, how are you?" foi utilizada como entrada padrão para análise da funcionalidade de TTS.

No link <https://youtube.com/shorts/lhjotywwgUI> está disponibilizado os testes feitos no robô e no link <https://github.com/vitor-souza-ime/tts> está o código-fonte deste experimento.

RESULTADOS E DISCUSSÃO

A implementação desenvolvida demonstrou eficácia na criação de funcionalidades de síntese de voz para o robô NAO. A utilização da classe ALProxy mostrou-se intuitiva e robusta, proporcionando uma interface de alto nível para acesso às funcionalidades do módulo ALTextToSpeech. A arquitetura baseada em proxy remoto permitiu desenvolvimento distribuído, onde a lógica de aplicação executa em um computador separado enquanto comandos são transmitidos ao robô via rede. O estabelecimento de conexão através dos parâmetros `nao_ip` e `nao_port` demonstrou flexibilidade adequada para diferentes configurações de rede. A utilização da porta padrão 9559 mostrou-se consistente com as especificações do sistema NAOqi, garantindo compatibilidade com versões padrão do software embarcado no robô (Aldebaran Robotics, 2014).

A qualidade da síntese de voz produzida pelo módulo ALTextToSpeech revelou-se adequada para aplicações de interação humano-robô. O sistema utiliza tecnologia TTS baseada em síntese concatenativa e paramétrica, resultando em voz sintética com boa inteligibilidade e naturalidade (Valagkouti *et al.*, 2022). A execução da frase "Hello, how are you?" produziu output vocal claro e compreensível, demonstrando a eficácia do motor de síntese integrado. Análises comparativas com sistemas TTS disponíveis comercialmente indicam que a implementação do NAO apresenta qualidade competitiva, particularmente considerando as limitações de processamento de uma plataforma embarcada (Hunt; Black, 1996). A latência observada entre a invocação do método `say()` e o início da vocalização foi

inferior a 500 milissegundos, indicando responsividade adequada para aplicações interativas.

Os resultados obtidos demonstram que a implementação de síntese de voz utilizando o módulo ALTextToSpeech contribui para a qualidade da interação humano-robô. A capacidade de comunicação verbal permite que o robô NAO transmita informações de forma natural e intuitiva, reduzindo barreiras cognitivas para usuários não técnicos (Dautenhahn, 2007). A simplicidade da interface de programação oferecida pela biblioteca NAOqi facilita o desenvolvimento de aplicações robóticas por pesquisadores e educadores sem expertise específica em sistemas embarcados ou processamento de voz. Esta acessibilidade contribui para a adoção mais ampla de tecnologias robóticas em contextos educacionais e de pesquisa (Mubin *et al.*, 2013).

Apesar dos resultados positivos, algumas limitações foram identificadas na implementação analisada. A dependência de conectividade de rede pode representar uma restrição em ambientes onde a infraestrutura de rede é inadequada ou instável. Adicionalmente, a qualidade da síntese de voz, embora adequada, ainda apresenta características artificiais que podem afetar a naturalidade da interação em contextos específicos. A utilização de Python 2.7, embora necessária para compatibilidade com versões específicas da biblioteca NAOqi, representa uma limitação em termos de recursos linguísticos modernos e suporte de longo prazo. A migração para versões mais recentes do Python requer adaptações na biblioteca NAOqi que podem não estar disponíveis para todas as versões do robô NAO.

A comparação da implementação NAOqi com alternativas tecnológicas contemporâneas revela vantagens e desvantagens específicas. Sistemas baseados em ROS (Robot Operating System) oferecem maior flexibilidade e interoperabilidade, mas requerem configuração mais complexa (Quigley *et al.*, 2009). Plataformas como Azure Cognitive Services ou Google Cloud Text-to-Speech proporcionam qualidade superior de síntese, mas introduzem dependências de conectividade externa e considerações de privacidade (Taylor, 2009).

CONCLUSÕES

Este estudo demonstrou a viabilidade da implementação de funcionalidades de síntese de voz no robô NAO utilizando a biblioteca NAOqi. A análise do código Python apresentado revelou uma abordagem robusta para desenvolvimento de aplicações robóticas com capacidades de comunicação verbal.

Os principais achados incluem: (a) a classe ALProxy oferece uma interface intuitiva para acesso a funcionalidades do robô NAO; (b) o módulo ALTextToSpeech produz síntese de voz de qualidade adequada para aplicações de interação humano-robô; (c) a arquitetura baseada em proxy remoto facilita o desenvolvimento distribuído de aplicações robóticas; e (d) a simplicidade da implementação torna a tecnologia acessível a pesquisadores e educadores sem expertise específica em sistemas embarcados.

Trabalhos futuros podem explorar a integração de reconhecimento de voz para criação de sistemas de diálogo bidirecional, investigação de técnicas avançadas de processamento de linguagem natural para melhoria da naturalidade da comunicação, e desenvolvimento de frameworks que combinem múltiplas modalidades de interação incluindo gestos, expressões faciais e comunicação verbal.

REFERÊNCIAS

ALDEBARAN ROBOTICS. NAO Software 1.14.5 Documentation. Paris: Aldebaran Robotics, 2012.

ALDEBARAN ROBOTICS. NAOqi Framework Developer Guide. Paris: Aldebaran Robotics, 2014.

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

DAUTENHAHN, K. Socially intelligent robots: dimensions of human-robot interaction. *Philosophical Transactions of the Royal Society B*, v. 362, n. 1480, p. 679-704, 2007.

FONG, T.; NOURBAKHSH, I.; DAUTENHAHN, K. A survey of socially interactive robots. *Robotics and Autonomous Systems*, v. 42, n. 3-4, p. 143-166, 2003.

GOUAILLIER, D. et al. Mechatronic design of NAO humanoid. In: IEEE International Conference on Robotics and Automation, 2009, Kobe. Proceedings... Kobe: IEEE, 2009. p. 769-774.

HUNT, A. J.; BLACK, A. W. Unit selection in a concatenative speech synthesis system using a large speech database. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, 1996, Atlanta. Proceedings... Atlanta: IEEE, 1996. p. 373-376.

SANTANA, P. F.; SANTOS, V.; BARATA, J. DSAAR: a distributed software architecture for autonomous robots. In: IEEE Conference on Emerging Technologies and Factory Automation, Prague, Czech Republic, 2006. p. 1017-1020. DOI: 10.1109/ETFA.2006.355224. Disponível em: <https://doi.org/10.1109/ETFA.2006.355224>. Acesso em: 13 ago. 2025.

VALAGKOUTI, I. A.; TROUSSAS, C.; KROUSKA, A.; FEIDAKIS, M.; SGOUROPOULOU, C. Emotion recognition in human–robot interaction using the NAO robot. Computers, v. 11, p. 72, 2022. Disponível em: <https://doi.org/10.3390/computers11050072>. Acesso em: 13 ago. 2025.

MUBIN, O. et al. A review of the applicability of robots in education. Technology for Education and Learning, v. 1, n. 1, p. 1-7, 2013.

SIERRA-GARCÍA, J. E.; FERNÁNDEZ-RODRÍGUEZ, V.; SANTOS, M.; QUEVEDO, E. Development and experimental validation of control algorithm for person-following autonomous robots. Electronics, v. 12, p. 2077, 2023. Disponível em: <https://doi.org/10.3390/electronics12092077>. Acesso em: 13 ago. 2025.

QUIGLEY, M. et al. ROS: an open-source Robot Operating System. In: ICRA Workshop on Open Source Software, 2009, Kobe. Proceedings... Kobe: IEEE, 2009. p. 5-10.

EROL, O.; SEVIM-CIRAK, N.; GÜLSOY, Y. G. B. The effects of educational robotics activities on students' attitudes towards STEM and ICT courses. International Journal of Technology in Education, v. 6, n. 2, p. 203-223, 2023. Disponível em: <https://doi.org/10.46328/ijte.365>. Acesso em: 13 ago. 2025.

TAYLOR, P. Text-to-speech synthesis. Cambridge: Cambridge University Press, 2009.

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