

Alexa for AFM, Microscopy Hackaton 2025

Hans Gunstheimer^{1,2}, Héctor Corte-León²

1) Institute of Microstructure Technology, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

2) Nanosurf AG, Liestal, Switzerland

E-mail: gunstheimer@nanosurf.com, corte@nanosurf.com

Abstract

Have you ever talked to your microscope? We present a shift in scanning probe microscopy (SPM) operation: The proof-of-concept for a voice-controller, AI-integrated framework for SPM operation. By bridging Speech-to-Text (STT) with instrument application programming interfaces (APIs) through large language models (LLMs), we enable hands-free execution of complex imaging workflows. This system is not merely a convenience for constrained environments like glove boxes, it is a foundational step towards autonomous robotic microscopy.

Introduction

Atomic Force Microscopy (AFM)^[1] is a fundamental technique for nanoscale characterization, yet its operation typically relies on complex graphical user interfaces (GUIs), requiring extensive training and limiting accessibility in high-throughput or remote environments. To overcome these constraints, a transition towards natural user interfaces (NUIs), such as voice-based control or touch and gesture interfaces is essential. Here, we present an approach integrating speech recognition to enable intuitive, hands-free interaction with AFM systems. Speech-to-text (STT) conversion reduces cognitive load and improves ergonomics during experimental workflows. Furthermore, coupling STT with large language models (LLMs) enables translation of voice commands into structured instrument control, parameter optimization, and real-time troubleshooting, bridging the gap between user intent and system execution. Previous studies have demonstrated the feasibility of LLM-driven AFM operation^[2,3]. This paradigm shift towards non-GUI interaction enhances usability, democratizes AFM access for non-expert users, and opens pathways for automation and adaptive experimentation. Our framework represents a first step in applying speech recognition to AFM control, paving the way for more natural and efficient human–instrument interaction.

Results

The goal of the challenge was to operate a scanning probe microscope by human voice commands. The system architecture (Figure 1) integrates a microphone for voice input, processed by a Python-based speech recognition module. The main loop continuously listens for input with a phrase time limit set to 5 s and a timeout of 10 s. Recognized commands are validated against a predefined look-up table and, if valid, trigger corresponding functions via the instrument application programming interface (API). A speaker provides optional a user feedback for confirmation. The SPM is then reacting accordingly to the called function. Here, we cover a basic workflow of an AFM measurement from laser alignment, approach, start and stop imaging, and withdraw. In Future, the look-up table can be replaced by an LLM to react to all kinds of vocal input, not just selected (and hard-coded)

commands. As microscope a stand-alone AFM (DriveAFM, Nanosurf AG)^[4] was used along with the instrument control software Nanosurf Studio (version 15.14) in combination with the Nanosurf Python API. For speech recognition we used a python speech recognition package. Easy-to-pronounce commands were selected to minimize word error rate. Error handling was implemented for unclear recognition or technical issues, with fallback feedback provided via audio and console messages. One command was defined to test the speech recognition, where just an audio feedback was provided without further instrument actions. Another command was defined to exit the listening loop. Overall, most of the commands were recognized, we had a pleasant user experience and were able to capture images without touching a keyboard or mouse.

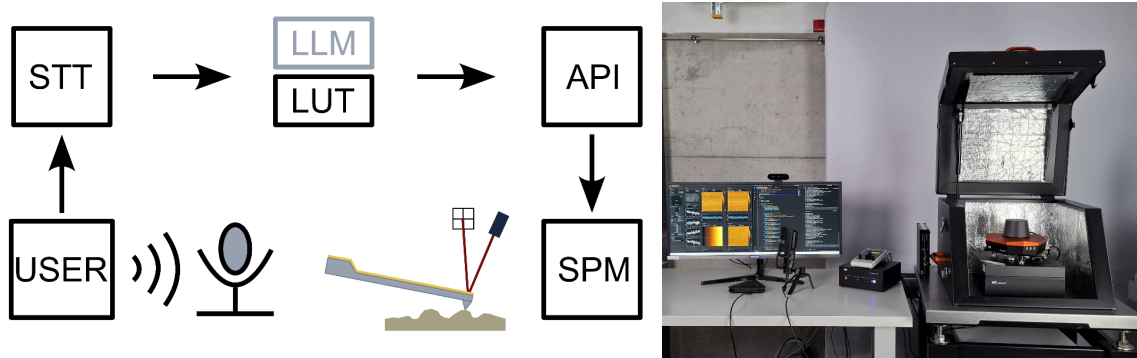


Figure1: Setup for controlling a scanning probe microscope (SPM) via voice commands. A speech-to-text (STT) module is recording the voice of a user, which is forwarded to either a look-up table (LUT) or a large language model (LLM). Based on this input, a function from the instrument application programming interface (API) is called and the SPM is executing the command.

Conclusion

This proof-of-concept demonstrates that voice-controlled AFM operation is both feasible and effective for basic imaging workflows. By eliminating the need for manual input, this approach improves ergonomics and accessibility. Future integration of LLMs with STT will enable flexible command interpretation, adaptive parameter optimization, and conversational troubleshooting, enabling fully autonomous and intelligent AFM systems.

References

- [1] G. Binnig, C. F. Quate, Ch. Gerber, *Phys. Rev. Lett.* **1986**, 56, 930.
- [2] Y. Liu, M. Checa, R. K. Vasudevan, *Mach. Learn. Sci. Technol.* **2024**, 5, 02LT01.
- [3] I. Mandal, J. Soni, M. Zaki, M. M. Smedskjaer, K. Wondraczek, L. Wondraczek, N. N. Gosvami, N. M. A. Krishnan, *Nat. Commun.* **2025**, 16, 9104.
- [4] J. D. Adams, P. L. T. M. Frederix, C. A. Bippes, *Rev. Sci. Instrum.* **2021**, 92, 129503.