

A Survey on Offline Voice-Controlled Humanoid Robotic Leg Systems Focusing on Kinematics, Embedded Control and Actuation Design

Dr. A. S. Shirkande¹; Kharche Vaishnavi Balaji²; Chavan Suhani Ashok³; Devkate Shubham Bhimrav⁴

^{1;2;3;4}Department of Electronics and Telecommunication Engineering SPVP's S.B. Patil College of Engineering Indapur, Maharashtra, India

Publication Date: 2026/10/13

Abstract: Humanoid and rehabilitation-oriented robotic legs are increasingly expected to respond to natural, hands-free user intent while remaining low-cost, private and safe. This survey reviews four recent works that, taken together, cover the main building blocks of a voice-commanded humanoid leg: (i) an offline, privacy-preserving voice-to-motion framework that couples Vosk speech recognition, tolerant command matching and forward-kinematics (FK) motion generation [1]; (ii) a servo-driven three-joint robotic leg with inverse-kinematics (IK) control and measured joint-angle accuracy [2]; (iii) a human-scale electric single-leg robot with an actuated toe and co-actuation, emphasising power electronics, actuator identification and balance control [3]; and (iv) a decoupled actuation humanoid leg evaluated with a new swing-time metric [4]. We compare their architectures, sensing, control and validation levels, identify gaps (notably the scarcity of hardware-validated, fully offline voice-controlled legs), and use the findings to motivate the design of a low-cost offline voice-controlled humanoid robotic leg. The system uses an ELECHOUSE V3 speaker-dependent voice recognition module and an ESP32 interface controller (voice-command relay, MPU6050 balance monitoring and a local Wi-Fi web application for manual control), an STM32 controller that handles the obstacle sensor and drives ten servo motors through a 16-channel PWM driver, and a 2S Liion/LiPo battery with separate regulated rails. Design considerations, a proposed command-to-motion mapping, safety logic and an evaluation plan are presented.

Keywords: Humanoid Robotic Leg, Offline Voice Control, Forward Kinematics, Inverse Kinematics, STM32, ESP32, MPU6050, Servo Actuation, Obstacle Avoidance, Survey.

How to Cite: Dr. A. S. Shirkande; Kharche Vaishnavi Balaji; Chavan Suhani Ashok; Devkate Shubham Bhimrav (2026) A Survey on Offline Voice-Controlled Humanoid Robotic Leg Systems Focusing on Kinematics, Embedded Control and Actuation Design. *International Journal of Innovative Science and Research Technology*, 11(10), 352-356. <https://doi.org/10.38124/ijisrt/26oct469>

I. INTRODUCTION

Legged robots that imitate the human lower limb are used for gait research, rehabilitation, education and as building blocks of humanoid platforms. Their usability depends on how naturally a person can instruct them. Voice is attractive because it is hands-free and accessible to users who cannot operate joysticks or wearable sensors. However, mainstream voice assistants rely on cloud speech recognition, which introduces network dependency, latency and privacy concerns [1]. For a robot that must react to words such as forward, backward, left, right and stop, an offline pipeline running on local, inexpensive hardware is a natural fit. Beyond the user interface, a robotic leg is a mechatronic system: its behaviour is shaped by kinematic modelling [1], [2], actuator selection and placement [3], [4], power delivery [3], sensing for balance and obstacle awareness, and the safety logic that decides when a spoken command may be executed

[1]. The literature usually treats these aspects separately. Our project, an offline voice-controlled humanoid robotic leg (two legs, five servos each), sits at their intersection.

Scope and method. This survey analyses four reference works in depth [1]–[4] and relates them to the selected hardware [8]–[15]. The works were read in full; the quantitative statements below are those reported by the respective authors, and we flag places where the reported figures are internally inconsistent or incomplete. The survey does not claim experimental results for our own system. Further background on sensors for robotics [5] and on rocker-bogie and six-legged locomotion platforms [6], [7] is available in the literature.

II. BACKGROUND

➤ A Kinematics for Leg Control

A planar leg is commonly modelled as serial links connected by revolute joints. In forward kinematics (FK) the joint angles are inputs and the link positions are computed directly; for a two-link thigh–shank model with hip at the origin [1]:

$$x_{knee} = l_1 \sin \theta_1, \ $$

$$y_{knee} = -l_1 \cos \theta_1$$

$$x_{ankle} = x_{knee} + l_2 \sin(\theta_1 + \theta_2), \quad \ $$

$$y_{ankle} = y_{knee} - l_2 \cos(\theta_1 + \theta_2)$$

FK is deterministic, cheap and free of singularities, which makes it suitable for real-time animation and for microcontrollers [1]. Inverse kinematics (IK) works the other way round: given a desired foot position it returns joint angles. For the three-joint (hip–knee–ankle) leg of [2] the knee angle follows from the cosine rule, $\theta_2 = \cos^{-1}((x^2 + y^2 - a_1^2 - a_2^2)/(2a_1a_2))$, followed by the hip and ankle angles. IK gives precise foot placement but can have multiple solutions and singularities [1], [2]. [4] shows that a closed-form IK can also be derived for a more complex decoupled leg, which keeps the controller simple.

➤ *Offline Speech Interfaces*

Offline recognisers (for example Vosk, used in [1]) run on laptops, phones or embedded boards without internet access, and so avoid cloud latency and upload of voice data. For a small, closed command vocabulary, dedicated voice-recognition modules or keyword-spotting firmware can replace a full speech-to-text engine; such modules typically return a command index rather than a transcript or confidence score, which affects the safety strategy discussed in Section V-E.

➤ *Servo-Driven Versus Torque-Controlled Legs*

Hobby/educational legs use position-controlled servos [2], where the controller only sends target angles and the servo closes the loop internally. Research legs such as [3] and [4] use proprioceptive, torque-controllable actuators with high-rate control (for example 500 Hz over CAN FD in [3]). The first is cheap and simple but offers limited torque, no torque feedback and coarse compliance; the second gives dynamic capability at much higher cost and complexity.

➤ *Motivatation*

Humanoid robotic legs have important applications in gait research, rehabilitation, education, and humanoid robotics. However, effective control of such systems requires the integration of voice interaction, kinematic modelling, actuation, sensing, balance monitoring, and safety mechanisms. Voice control is particularly attractive because it provides a hands-free and accessible method for interacting with robotic systems.

Existing studies address different parts of this problem. Some focus on offline voice-to-motion control and kinematic modelling, while others concentrate on servo-driven robotic legs, actuator design, balance control, and humanoid locomotion. However, these aspects are often studied separately, and there is limited work combining offline voice control, obstacle sensing, inertial feedback, and multi-servo humanoid leg actuation on a low-cost embedded platform.

Therefore, this survey is motivated by the need to systematically study and compare existing approaches, identify their strengths and limitations, and determine the research gap for developing a low-cost, offline voice-controlled humanoid robotic leg. The findings of this survey provide a foundation for selecting suitable control, sensing, actuation, and safety approaches for the proposed system.

➤ *Comparative Analysis*

Table 1 summarises the four works along the dimensions most relevant to a voice-controlled humanoid leg.

Table 1 Side by Side Comparison of the Four Reference Works.

Aspect	[1] Offline voice LLRR	[2] Motion-capture robotic leg	[3] StaccaToe	[4] DecARt Leg
Main focus	Offline voice-to-motion pipeline with tolerant matching and confidence gating	Servo robotic leg replaying captured human leg swing	Human-scale single leg with actuated toe, co-actuation	Decoupled-actuation humanoid leg and FAST metric
Structure / DOF	2-link planar leg (hip, knee), simulated	3 joints (hip, knee, ankle), 3 DOF, planar	6 actuators incl. 2-DoF ankle and toe	Quasi-telescopic leg, 2-DoF ankle, all motors above knee
Actuation	None (simulation)	DC/RC servos	Electric proprioceptive actuators (AK series), 48 V	Rotational proprioceptive actuators
Controller / compute	MATLAB GUI + Python (Vosk) on a PC	PC + ATmega328 + servo controller	UP Xtreme i7, MAB MD80 drivers, 500 Hz CAN FD	PC/embedded; PyBullet/MuJoCo, TSID, ALIPM, RL
Kinematics	FK (Eqs. 1–2)	DH-based IK (cosine	Single rigid-body	Closed-form IK

		rule)	trajectory optimisation, WBIC	
Sensing	Microphone only (IMU as future work)	Camera and markers	Onboard IMU, motor encoders	Joint encoders, IMU (humanoid)
Voice / HRI	Yes: Vosk, 5 commands, offline	No	No	No
Safety logic	Confidence-gated execution; E-stop in roadmap	Not discussed	Torque/voltage limits; planar constraint in prior work	Joint limits, simulation checks
Validation	Simulation only	Hardware prototype	Hardware: tip-toe balance, jump	Simulation + preliminary hardware
Key reported result	80–95% accuracy, 0.89–1.32 s latency	Mean angle error 1.46° / 1.66° / 0.46°	Knee torque > 100 N·m via co-actuation	FAST 0.17 s; 2.2 m/s sim. walking
Main limitation	No hardware; incomplete noise results	Offline playback; dynamics ignored; no user-intent input	Single leg, no hip-body locomotion; voltage sag limits jump	Complex mechanism; limited hardware data
Use for our project	Voice pipeline, safety gating, FK, test plan	Servo leg hardware, IK, error metrics	Power, backlash, cabling, balance lessons	Actuator placement, closed-form IK, metric

III. PROPOSED OFFLINE VOICE-CONTROLLED HUMANOID LEG SYSTEM

➤ Architecture

Figure 1 shows the proposed architecture. The ESP32 is the communication and interface controller: it receives recognised commands from the voice module over UART, serves a local Wi-Fi web application with Forward, Backward, Left, Right and Stop controls, reads the MPU6050 over I2C, maps voice and web inputs to movement commands and sends them to the STM32 over UART. The STM32 is the actuation controller: it processes the obstacle sensor and drives the servo driver, which drives five servos

per leg (ten in total). Keyword spotting is performed entirely inside the voice module, not on the ESP32, and Bluetooth is not used. Power comes from a 2S battery through a high-current 6 V servo rail and a separate regulated rail for the controllers. “Offline” here means that no internet connection or cloud service is needed; the web application is served locally by the ESP32.

➤ Hardware Components Required

Table 2 lists the selected components and their roles. Specifications are summarised from the supplied component references [8]–[15] and should be verified against official datasheets before final submission.

Table 2 Component Summary

Component	Role in system	Key characteristics (from references / project data)	Ref.
STM32 (F103C8T6)	Actuation controller: receives movement commands from the ESP32 (UART), reads the IR obstacle sensor, generates joint set-points and commands the servo driver (I2C)	32-bit Arm Cortex-M3, up to 72 MHz, 64 KB flash, 20 KB SRAM, I2C/UART/SPI, 3.3 V logic	[8]
ESP32	Communication and interface controller: receives voice commands (UART), hosts the local Wi-Fi web app, reads the MPU6050 (I2C), maps voice/web inputs to movement commands and sends them to the STM32 (UART)	Dual-core Xtensa, up to 240 MHz, integrated Wi-Fi (Bluetooth present but not used), 3.3 V logic	[9]
Voice recognition module (ELECHOUSE V3)	Keyword spotting of trained commands (forward, backward, left, right, stop)	Speaker-dependent; up to 80 stored commands, 7 active at a time; command length up to about 1500 ms	[12]
IR obstacle sensor	Detects obstacles in the walking direction (read by the STM32); triggers stop	IR emitter–receiver pair, digital output, adjustable sensing range (short range, typically a few cm to tens of cm)	[14]
MPU6050	Tilt/balance monitoring (roll, pitch), read by the ESP32 over I2C	3-axis gyroscope + 3-axis accelerometer, 16-bit ADCs, I2C, optional on-chip motion processing	[15]
16-channel PWM servo driver (PCA9685)	Generates PWM for 10 servos over I2C (from the STM32)	16 channels, 12-bit resolution, I2C, separate servo supply (V+ up to 6 V on the Adafruit board)	[10]
MG90S servo (×10)	Joint actuators (5 per leg)	Metal-gear micro servo, 4.8–6 V, stall torque about 1.8 kg·cm (4.8	[11]

		V) to 2.2 kg·cm (6 V), about 0.1 s/60°, 13.4 g, ~180° range	
Battery and power rails	2S Li-ion/LiPo pack feeding a high-current BEC/buck converter (servo rail, about 6 V) and a separate regulator path for the ESP32 and STM32	7.4 V nominal, 8.4 V fully charged, about 4000–5000 mAh; BEC/buck rating to be chosen from the servo peak current; the voice module needs 4.5–5.5 V	[13]

➤ Command-to-Motion Mapping (Proposed)

Table 3 Proposed Mapping Between Spoken Commands and Motion Primitives.

Voice command	Intended behaviour	Safety / notes
Forward	Start walking-gait cycle forward (alternating swing and stance of the two legs)	Blocked or interrupted if the IR sensor reports an obstacle or tilt exceeds the limit
Backward	Reverse gait cycle	No obstacle sensor coverage at the rear unless added; use reduced speed
Left / Right	Turn by hip rotation (if available) or by asymmetric step pattern	Joint allocation of the five servos per leg to be confirmed
Stop	Return smoothly to a stable neutral standing pose and hold	Highest priority; accepted at any time, as [1] treats stop/rest as returning to neutral

➤ Safety and Reliability

[1] gates execution on match similarity and ASR confidence. The V3 module is speaker-dependent and provides no numerical confidence: it reports the index of a trained command when it matches and otherwise reports nothing, so the confidence gating of [1] cannot be copied directly. The following measures compensate: (i) train every command carefully for each intended user and keep the vocabulary small (five commands; the module allows seven active at a time, so one slot could hold an arming word); (ii) the ESP32 acts as arbiter between the voice and web sources, accepts motion commands only in a valid state and accepts stop from either source immediately; (iii) a heartbeat/time-out on the ESP32–STM32 UART link makes the STM32 return the leg to a safe neutral pose if no valid frame arrives, which also covers an ESP32 crash or Wi-Fi loss; (iv) the obstacle interlock is enforced locally by the STM32, independently of the ESP32; (v) the ESP32 estimates tilt from the MPU6050 and sends a flag, on which the STM32 freezes the gait and holds a safe pose; (vi) the web application is protected by a Wi-Fi password and kept on a local network; (vii) a physical emergency switch cuts the servo rail, in line with the E-stop and dead-man enable proposed in [1]; and (viii) the system is first tested on a bench with the body supported, following the staged roadmap in [1].

IV. FUTURE SCOPE

- IMU-gated gait timing (stance/swing phase detection) so that commands take effect at safe instants, as proposed in [1].
- Closed-form IK for foot-trajectory generation and a calibrated kinematic model of the built leg, following [2] and [4].
- Larger vocabulary and speaker-independent or multi-user recognition (the V3 module is speaker-dependent), and secure local web access; the ESP32 could also provide telemetry.

- Servo upgrade (higher torque or feedback-capable actuators), better power distribution and, later, torque-controlled joints [3], [4].
- Clinical-oriented extensions (rehabilitation exercises, supervised pilots) only after hardware safety validation.

➤ Proposed Evaluation Plan

To make the project results comparable with the reviewed literature, we propose the following metrics: command recognition accuracy per command and per noise condition with Wilson confidence intervals, and false-positive rate from no-command/noise-only trials, as defined in [1]; end-to-end latency from utterance to first joint motion and to completed motion (mean $\pm$ SD, $n \geq 20$, rather than $n = 5$); joint-angle tracking error between commanded and measured angles (as in [2]); tilt stability from MPU6050 during stand and walk; obstacle stop distance and stop time; emergency stop and time-out response; and battery run time and supply-voltage sag during walking (motivated by [3]). Results should be reported separately for quiet and noisy conditions and per trained user; because the module is speaker-dependent, also test whether untrained speakers are rejected rather than mis-recognised.

V. CONCLUSION

The reviewed works cover complementary parts of the problem: [1] supplies an offline voice pipeline and a confidence-gated safety policy but stops at simulation; [2] demonstrates an affordable servo-driven leg with measured joint accuracy; [3] shows how power electronics, backlash, cabling and balance control determine real-world performance; and [4] shows how actuator placement and a closed-form kinematic model improve agility and simplify control. No reviewed system combines offline voice commands, obstacle sensing, inertial feedback and a multi-servo humanoid leg on low-cost embedded hardware. The proposed STM32/ESP32-based system, with its priority-

based safety logic and defined evaluation plan, targets exactly this gap.

REFERENCES

- [1]. N. S. M. Ali, M. H. Saleh, and N. H. Abbas, "Offline voice-controlled lower limb rehabilitation robot using forward kinematics and fuzzy command matching," *IEEE Access*, vol. 14, pp. 84453–84465, 2026, doi: 10.1109/ACCESS.2026.3698788.
- [2]. R. T. Yunardi, A. A. Firdaus, and E. I. Agustin, "Robotic leg design to analysis the human leg swing from motion capture," *Bulletin of Electrical Engineering and Informatics*, vol. 6, no. 3, pp. 256–264, Sep. 2017, doi: 10.11591/eei.v6i3.645.
- [3]. N. Perera, S. Yu, D. Marew, M. Tang, K. Suzuki, A. McCormack, S. Zhu, Y.-J. Kim, and D. Kim, "StaccaToe: A single-leg robot that mimics the human leg and toe," *arXiv:2404.05039v1 [cs.RO]*, Apr. 2024.
- [4]. E. Davydenko, A. Volchenkov, V. Gerasimov, and R. Gorbachev, "DecARt Leg: Design and evaluation of a novel humanoid robot leg with decoupled actuation for agile locomotion," *arXiv:2511.10021v1 [cs.RO]*, Nov. 2025.
- [5]. Bodhale, Umesh, "Overview of Sensors for Robotics," K. Aparna, *International Journal of Engineering Research and Technology (IJERT)*, vol. 2. [Online]. Available: <https://www.ijert.org/research/overview-of-sensors-for-robotics-IJERTV2IS3450.pdf>
- [6]. Pranav Patil, Prof. Aparna Shirkande, Suraj Magar, and Prathamesh Lokhande, "Rocker-Boogie for Military Use," *International Journal of Scientific Research in Engineering and Management (IJSREM)*, vol. 6, no. 11, p. 5, Nov. 2022. [Online]. Available: www.ijisrem.com
- [7]. S. Kale Aparna and D. Salunke Geeta, "Six legged locomotion on uneven terrain." [Online]. Available: https://scholar.google.com/citations?view_op=view_citation&hl=en&user=v07QP-UAAAAJ&citation_for_view=v07QP-UAAAAJ:W7OEmFMyl1HYC
- [8]. "STM32F103C8T6 datasheet explained," *Ultra Librarian*, May 2026. [Online]. Available: <https://www.ultralibrarian.com/2026/05/28/stm32f103c8t6-datasheet-explained-ulc>
- [9]. ESP32 specification summary (link supplied by the project team). [Online]. Available: <https://share.google/aimode/TBXTDLO7nSIumPUW X> — replace with the official Espressif ESP32 datasheet.
- [10]. Adafruit, "Adafruit 16-channel 12-bit PWM/servo driver," *Adafruit Learning System*. [Online]. Available: <https://cdn-learn.adafruit.com/downloads/pdf/16-channel-pwm-servo-driver.pdf>
- [11]. Components101, "MG90S metal gear servo motor." [Online]. Available: <https://components101.com/motors/mg90s-metal-gear-servo-motor>
- [12]. ELECHOUSE, "Voice Recognition Module V3 (VR3)," specifications as supplied by the project team. [Online]. Available: <https://share.google/aimode/9Cz5uKvvCXkTBIdm8> — replace with the manufacturer user manual.
- [13]. Battery specification (link supplied by the project team). [Online]. Available: <https://share.google/aimode/5m1UWeoghiNZE47C4> — replace with the battery datasheet.
- [14]. IC-Components, "Infrared sensors: specs, operation, types and applications." [Online]. Available: <https://www.ic-components.com/blog/Infrared-Sensors-Specs,Operation,Types-and-Applications.jsp>
- [15]. MPU6050 specification summary (link supplied by the project team). [Online]. Available: <https://share.google/aimode/hWoKijFJ9uReaipyO> — replace with the InvenSense/TDK MPU-6050 datasheet.