

협동 용접 로봇 제어를 위한 LLM 기반 음성 제어 시스템

Taejoo Um^a, Minjoo Choi^a, Jaekyeong Lee^a and Woosung Kim^b

^aDivision of Naval Architecture and Ocean Systems Engineering, Korea Maritime and Ocean University, Busan, Korea

^bDigital Manufacturing Innovation Division, Research Institute of Medium & Small Shipbuilding, Busan, Korea

BACKGROUND / OBJECTIVES

산업용 용접 로봇은 팬던트 컨트롤러에 의존하고 있으며, 이는 열악한 작업 환경에서 파손·분실 위험이 크고 손을 이용한 조작의 불편함이 뒤따른다. 본 연구에서는 이러한 문제를 완화하기 위한 하나의 시도로, 오픈소스 기반의 대규모 언어 모델(LLM)을 활용하여 자연어 음성 명령으로 로봇을 제어하는 핸즈프리 음성 인터페이스의 가능성을 탐색하였다.

제안된 시스템은 기본적인 명령 인식과 제어가 가능함을 보여주었으며, 팬던트 컨트롤러에 대한 의존을 줄이고 작업 효율성 향상에 기여할 수 있음을 시사했다. 본 연구는 음성인식을 통한 로봇 제어 방식이 실제 산업 환경에서 실용적으로 활용될 수 있는지에 대한 기초적인 가능성을 검토하였다. 이러한 접근은 용접 작업 등 복잡하고 반복적인 작업 환경에서 음성 기반 인간-로봇 협업의 실현 가능성을 높이고, 작업 효율성과 안전성을 향상시키는 기반 기술로 발전할 수 있을 것으로 기대된다.



▲ 기존의 용접 로봇 제어 방식

METHODS

1. 환경 세팅

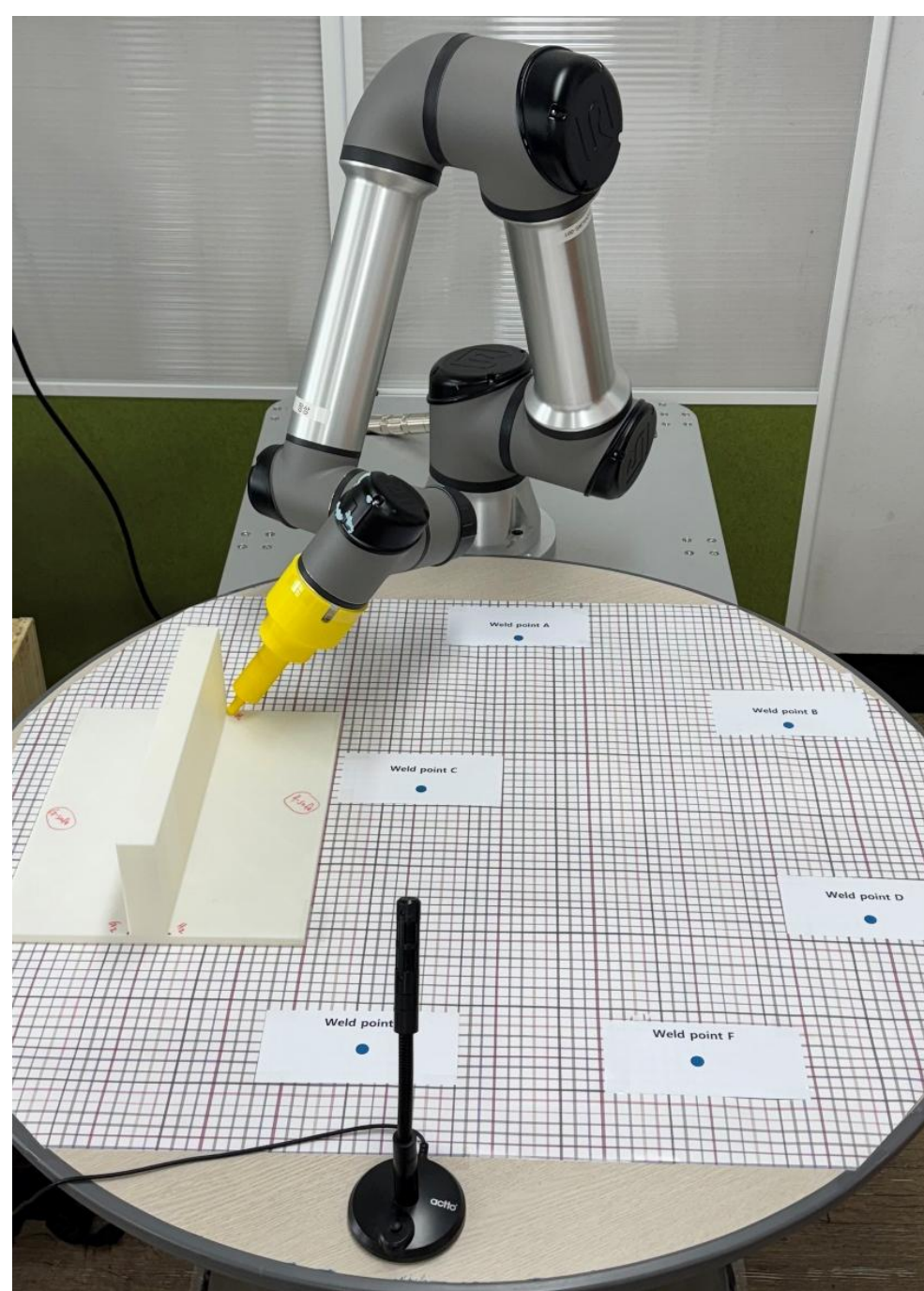
UR5e

Specification

Payload	5 kg (11 lbs)
Reach	850 mm (33.5 in)
Degrees of freedom	6 rotating joints
Programming	12 inch touchscreen with PolyScope graphical user interface
Power consumption (average)	570 W
Maximum power	200 W
Moderate operating settings	200 W
Operating temperature range	Ambient temperature: 0-50°C (32-122°F)
Safety functions	17 configurable safety functions
In compliance with	EN ISO 13849-1 (PLd Category 3) and EN ISO 10218-1

Physical

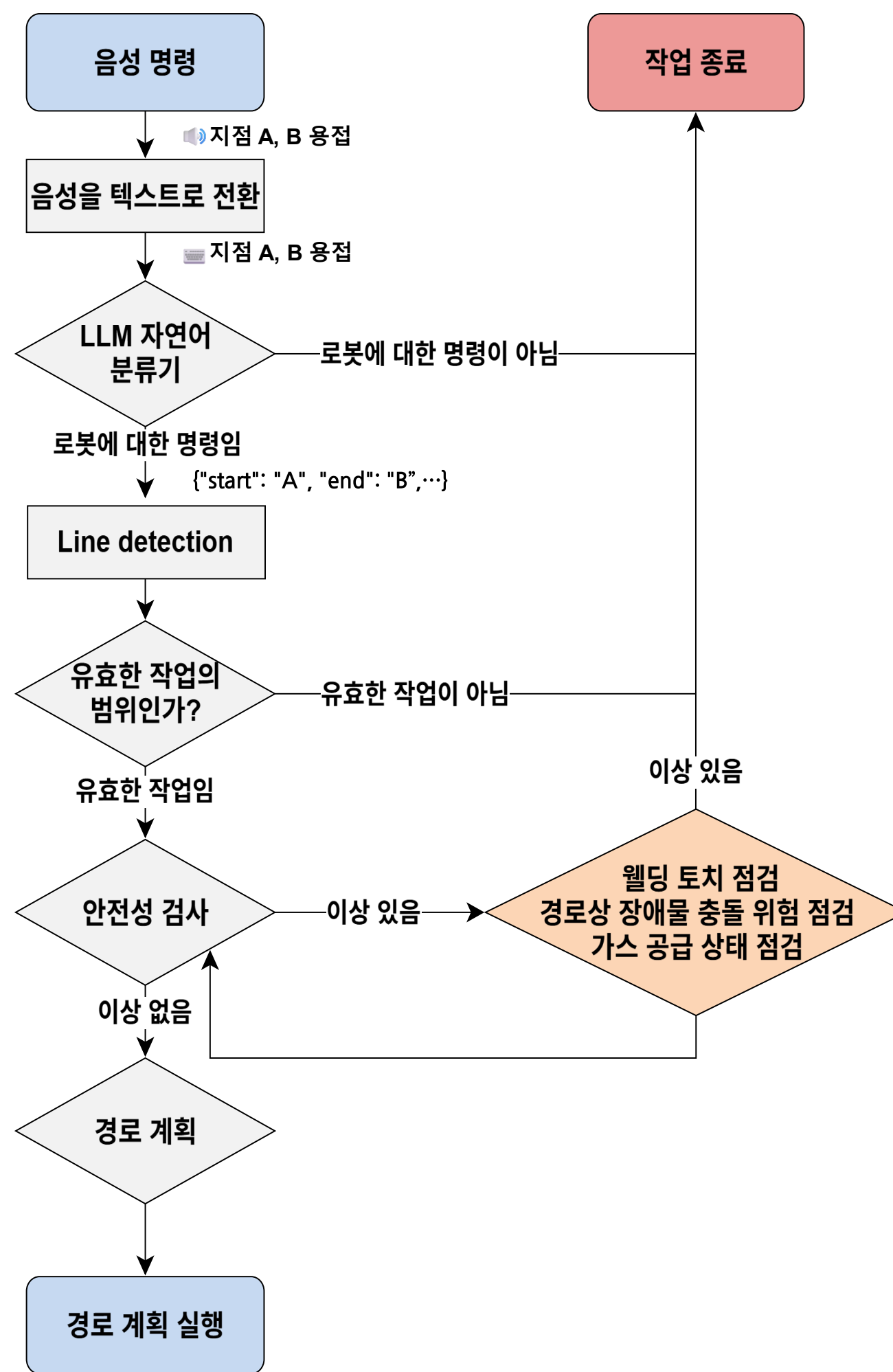
Footprint	Ø 149 mm
Materials	Aluminum, plastic, steel
Tool Flange	EN ISO-9409-1-50-4-M6
Connector type	M8 M8 8-pin
Cable length (robot arm)	6 m (236 in)
	12 m (472 in) and high-flex options available.
Weight including cable	20.6 kg (45.4 lbs)
Humidity	≤ 90% RH (non-condensing)



통제된 환경에서 작업할 수 있도록 바닥에 포인트를 고정시켰다. 각각 지점은 A~H2까지 존재하고 각 포인트의 TCP위치와 Joint 각도는 측정을 하여 코드상에 구현했다.

이 시스템에서는 총 2가지 용접 방식을 지원한다. 단순히 두 지점을 연결하는 직선용접, 두 지점을 지그재그로 용접하는 위빙용접이다.

2. Flowchart



3. Fine-tuning

Fine-tuning 목표 : 자연어 음성 명령을 로봇이 이해할 수 있는 정형화된 JSON 데이터로 정확하게 변환하는 것

기반 모델 : Whisper small, unsloth/Llama-3.2-3B-Instruct

핵심 기술 : LoRA(Low-Rank Adaptation)를 활용한 파라미터 효율적 미세조정

Rank	16
Lora alpha	16
Target modules	Attention & FFN Layers

Fine-tuning을 위한 데이터 예시

- {"text": "지점 A와 B 위빙으로 연결", "parse": {"start": "A", "end": "B", "type": "위빙용접", "speed": "0.05", "Hz": "2", "width": "8"}}
- {"text": "C로 이동해줘", "parse": {"start": "C", "end": "C", "type": "이동", "speed": "0.1", "Hz": "1", "width": "4"}}
- {"text": "B와 D를 직선용접. 속도 0.19.", "parse": {"start": "B", "end": "D", "type": "직선용접", "speed": "0.19", "Hz": "1", "width": "6"}}
- {"text": "C와 A를 위빙용접. 속도 0.12, 폭 9mm, 5Hz.", "parse": {"start": "C", "end": "A", "type": "위빙용접", "speed": "0.12", "Hz": "5", "width": "9"}}

RESULTS

No.	입력 문장 (Input Command)	모델 출력 (Parsed JSON)	정확도
1	지점 A와 B를 직선 용접으로 연결해줘	{"start": "A", "end": "B", "type": "직선용접", "speed": "0.05", "Hz": "2", "width": "8"}	0.00%
2	지점 B로 가	{"start": "B", "end": "B", "type": "이동", "speed": "0.1", "Hz": "1", "width": "4"}	0.00%
3	A에서 C로 위빙 용접 부탁해	{"start": "A", "end": "C", "type": "위빙용접", "speed": "0.05", "Hz": "2", "width": "8"}	100.00%
4	B와 C를 직선 용접으로 해줘	{"start": "B", "end": "C", "type": "직선용접", "speed": "0.19", "Hz": "1", "width": "6"}	100.00%
5	지점 A로 이동해	{"start": "A", "end": "A", "type": "이동", "speed": "0.1", "Hz": "1", "width": "4"}	83.30%
6	B와 A를 위빙 용접으로 연결	{"start": "B", "end": "A", "type": "위빙용접", "speed": "0.07", "Hz": "2", "width": "9"}	100.00%
7	C에서 A로 직선 용접	{"start": "C", "end": "A", "type": "직선용접", "speed": "0.15", "Hz": "2", "width": "4"}	100.00%
8	지점 C로 가봐	{"start": "C", "end": "C", "type": "이동", "speed": "0.1", "Hz": "1", "width": "4"}	0.00%
9	A와 B 직선 용접해	{"start": "A", "end": "B", "type": "직선용접", "speed": "0.12", "Hz": "5", "width": "9"}	83.30%
10	B에서 C로 이동	{"start": "B", "end": "C", "type": "이동", "speed": "0.1", "Hz": "1", "width": "4"}	0.00%
11	F 지점으로 이동합니다	{"start": "F", "end": "F", "type": "이동", "speed": "0.07", "Hz": "2", "width": "9"}	0.00%
12	E에서 F로 위빙용접. 속도 0.07, 폭 9mm, 2Hz.	{"start": "E", "end": "F", "type": "위빙용접", "speed": "0.07", "Hz": "2", "width": "9"}	83.30%
13	C 지점 위빙용접. 속도 0.15, 폭 4mm, 2Hz.	{"start": "C", "end": "C", "type": "위빙용접", "speed": "0.15", "Hz": "2", "width": "4"}	66.70%
14	H2 지점으로 이동.	{"start": "H2", "end": "H2", "type": "이동", "speed": "0.1", "Hz": "1", "width": "4"}	0.00%
15	A에서 G2로 위빙용접. 속도 0.07, 폭 9mm, 8Hz.	{"start": "A", "end": "G2", "type": "위빙용접", "speed": "0.07", "Hz": "8", "width": "9"}	100.00%
16	A와 D 지점 위빙용접. 속도 0.13, 폭 6mm, 1Hz.	{"start": "A", "end": "D", "type": "위빙용접", "speed": "0.13", "Hz": "1", "width": "6"}	100.00%
17	C와 H2를 위빙용접. 속도 0.10, 폭 7mm, 4Hz.	{"start": "C", "end": "H2", "type": "위빙용접", "speed": "0.10", "Hz": "4", "width": "7"}	100.00%
18	G1에서 G2로 위빙용접. 속도 0.07, 폭 9mm, 8Hz.	{"start": "G1", "end": "G2", "type": "위빙용접", "speed": "0.07", "Hz": "8", "width": "9"}	100.00%
19	D 지점으로 가세요	{"start": "D", "end": "D", "type": "이동", "speed": "0.1", "Hz": "1", "width": "4"}	0.00%
20	C 지점 위빙용접. 속도 0.14, 폭 5mm, 3Hz.	{"start": "C", "end": "C", "type": "위빙용접", "speed": "0.14", "Hz": "3", "width": "5"}	66.70%

▲ 기반 모델 추론 성능은 58.33% 정확도

No.	입력 문장 (Input Command)	모델 출력 (Parsed JSON)	정확도 (%)
1	지점 A와 B를 직선 용접으로 연결해줘	{"start": "A", "end": "B", "type": "직선용접", "speed": "0.05", "Hz": "2", "width": "8"}	100.00%
2	지점 B로 가	{"start": "B", "end": "B", "type": "이동", "speed": "0.1", "Hz": "1", "width": "4"}	100.00%
3	A에서 C로 위빙 용접 부탁해	{"start": "A", "end": "C", "type": "위빙용접", "speed": "0.05", "Hz": "2", "width": "8"}	100.00%
4	B와 C를 직선 용접으로 해줘	{"start": "B", "end": "C", "type": "직선용접", "speed": "0.19", "Hz": "1", "width": "6"}	100.00%
5	지점 A로 이동해	{"start": "A", "end": "A", "type": "이동", "speed": "0.1", "Hz": "1", "width": "4"}	100.00%
6	B와 A를 위빙 용접으로 연결	{"start": "B", "end": "A", "type": "위빙용접", "speed": "0.07", "Hz": "2", "width": "9"}	100.00%
7	C에서 A로 직선 용접	{"start": "C", "end": "A", "type": "직선용접", "speed": "0.15", "Hz": "2", "width": "4"}	100.00%
8	지점 C로 가봐	{"start": "C", "end": "C", "type": "이동", "speed": "0.1", "Hz": "1", "width": "4"}	100.00%
9	A와 B 직선 용접해	{"start": "A", "end": "B", "type": "직선용접", "speed": "0.12", "Hz": "5", "width": "9"}	100.00%
10	B에서 C로 이동	{"start": "B", "end": "C", "type": "이동", "speed": "0.1", "Hz": "1", "width": "4"}	100.00%
11	F 지점으로 이동합니다	{"start": "F", "end": "F", "type": "이동", "speed": "0.07", "Hz": "2", "width": "9"}	100.00%
12	E에서 F로 위빙용접. 속도 0.07, 폭 9mm, 2Hz.	{"start": "E", "end": "F", "type": "위빙용접", "speed": "0.07", "Hz": "2", "width": "9"}	100.00%
13	C 지점 위빙용접. 속도 0.15, 폭 4mm, 2Hz.	{"start": "C", "end": "C", "type": "위빙용접", "speed": "0.15", "Hz": "2", "width": "4"}	100.00%
14	H2 지점으로 이동.	{"start": "H2", "end": "H2", "type": "이동", "speed": "0.1", "Hz": "1", "width": "4"}	100.00%
15	A에서 G2로 위빙용접. 속도 0.07, 폭 9mm, 8Hz.	{"start": "A", "end": "G2", "type": "위빙용접", "speed": "0.07", "Hz": "8", "width": "9"}	100.00%
16	A와 D 지점 위빙용접. 속도 0.13, 폭 6mm, 1Hz.	{"start": "A", "end": "D", "type": "위빙용접", "speed": "0.13", "Hz": "1", "width": "6"}	100.00%
17	C와 H2를 위빙용접. 속도 0.10, 폭 7mm, 4Hz.	{"start": "C", "end": "H2", "type": "위빙용접", "speed": "0.10", "Hz": "4", "width": "7"}	100.00%
18	G1에서 G2로 위빙용접. 속도 0.07, 폭 9mm, 8Hz.	{"start": "G1", "end": "G2", "type": "위빙용접", "speed": "0.07", "Hz": "8", "width": "9"}	100.00%
19	D 지점으로 가세요	{"start": "D", "end": "D", "type": "이동", "speed": "0.1", "Hz": "1", "width": "4"}	100.00%
20	C 지점 위빙용접. 속도 0.14, 폭 5mm, 3Hz.	{"start": "C", "end": "C", "type": "위빙용접", "speed": "0.14", "Hz": "3", "width": "5"}	100.00%

▲ 기반 모델 Fine-tuning 이후 추론 성능은 100% 정확도

CONCLUSIONS

본 연구에서는 Whisper-small 음성인식 모델과 LLaMA 3.2 3B를 활용한 음성 제어 용접로봇 시스템을 성공적으로 개발하였다. 자연어 명령을 실시간으로 용접 작업으로 변환하는 자동화 시스템을 구현하여 기존 팬던트 사용 방식을 대체하는 직관적인 인간-로봇 상호작용 패러다임을 제시하였다. 향후 3D 비전 시스템 통합과 다중 로봇 협업으로 확장하여 산업 현장 적용성을 높일 계획이다.

REFERENCES.

1. Pires, J. Norberto, "Robot-by-voice: experiments on commanding an industrial robot using the human voice," Industrial Robot: An International Journal, vol. 32, no. 6, pp. 505-511, 2005.
2. Maksymova, Svitlana, Rami Matarneh, Vyacheslav V. Lyashenko, and Nataliya V. Belova, "Voice Control for an Industrial Robot as a Combination of Various Robotic Assembly Process Models," Journal of Computer and Communications, vol. 5, no. 11, pp. 1-15, 2017.

실제 시스템 구현

1. 음성 입력
 - ▶ 지점 A와 F를 속도 0.05rad/s, 폭 8mm, 주기 2Hz로 위빙용접 시작(음성)
2. 음성 → 텍스트 변환 (Whisper모델)
 - ▶ 지점 A와 F를 속도 0.05rad/s, 폭 8mm, 주기 2Hz로 위빙용접 시작(텍스트)
3. 입력된 텍스트가 용접로봇에 대한 명령인지 판단 (LLaMA)
 - ▶ 용접 지시로 분류되었습니다.
4. 입력된 텍스트를 JSON으로 변환 (LLaMA)
 - ▶ {"start": "A", "end": "F", "type": "위빙용접", "speed": "0.05", "Hz": "2", "width": "8"}

전체 시스템 응답 성능

구분	정확도
음성 인식	92.72%
용접 명령 분류	95%
텍스트→JSON 변환 (fine-tuning 전)	58.33%
텍스트→JSON 변환 (fine-tuning 후)	100%

No	구분	시간
1	모델 로드 (초기 1회)	36.0초
2	음성 인식 (Whisper)	0.02초
3	용접 명령 분류 (LLaMA)	0.007초
4	JSON 변환 (LLaMA)	0.024초
	총 응답시간 (로드 제외)	0.51초