

Speech-Based Human-Exoskeleton Interaction for Lower Limb Motion Planning

Eddie Guo
Cumming School of Medicine
University of Calgary
Alberta T2N 4N1, Canada
eddie.guo@ucalgary.ca

Christopher Perlette
Department of Neuroscience
University of Lethbridge
Alberta T2G 4V1, Canada
chris.perlette@uleth.ca

Mojtaba Sharifi
Department of Mechanical Engineering
San Jose State University
California 95192, USA
mojtaba.sharifi@sjsu.ca

Lukas Grasse
Department of Neuroscience
University of Lethbridge
Alberta T2G 4V1, Canada
lukas.grasse@uleth.ca

Matthew Tata
Department of Neuroscience
University of Lethbridge
Alberta T2G 4V1, Canada
matthew.tata@uleth.ca

Vivian K. Mushahwar
Department of Medicine
University of Alberta
Alberta T6G 2H5, Canada
vmushahw@ualberta.ca

Mahdi Tavakoli
Department of Elec & Comp Eng
University of Alberta
Alberta T6G 2R3, Canada
mahdi.tavakoli@ualberta.ca

Abstract—This study presents a speech-based motion planning strategy (SBMP) developed for lower limb exoskeletons to facilitate safe and compliant human-robot interaction. A speech processing system, finite state machine (FSM), and central pattern generator (CPG) are the building blocks of the proposed strategy for online planning of the exoskeleton’s trajectory. A novel set of CPG dynamics is proposed to synchronize time-continuous transitions between exoskeleton locomotion states (e.g., sit, stand, walk) in response to discrete user inputs, while speech inputs are processed through an FSM. According to experimental evaluations, this speech-processing system achieved low levels of word and intent errors. Regarding locomotion, the completion time for users with voice commands was 54% faster than that using a mobile app interface. With the proposed SBMP, users are able to maintain their postural stability with both hands free. This supports its use as an effective motion planning method for the assistance and rehabilitation of individuals with lower-limb impairments.

I. INTRODUCTION

Assistive robotic systems can enhance the quality of life of people affected by neurological impairments [1]. These systems include lower limb exoskeletons, such as ReWalk [2], Indego [3], HAL [4], and Exo-H3 (used in these experiments) [5], which are designed to rehabilitate individuals with neurological impairments. In comparison with traditional physical therapies, wearable exoskeletons allow users to interact more easily with their environment, improving mobility and independence in non-ambulatory individuals [1], [6], [7], [8]. To facilitate safe and compliant human-robot interactions (HRIs), the exoskeleton motion planning strategy should be intuitive and efficient to use [9]. Social HRI, where humans use body language, gestures, and speech to interact with robots, shows promise in addressing issues with physical interactions through quick and efficient identification of user intentions [10].

Although speech recognition (SR) is increasingly adopted for HRI, this adoption is generally limited to humanoid robots [10], [11]. Nonetheless, the integration of SR into other forms of robotics may improve their ergonomics and practicality for human use. For example, SR reduces the need for button-based control interfaces, allowing for hands-free use of an exoskeleton, benefitting users who require both hands to grasp gait aids. However, these interactions are only effective if the robot can perceive what a user is saying.

Computational SR is a task that requires turning auditory information into text. Over the last 40 years, there have been several technological advancements permitting accurate speech transcription even under adverse conditions [12]. In particular, a neural network architecture called the time delay neural network factorization model has been shown to be effective at acoustic modeling and speech perception [13], [14]. Natural language understanding (NLU) addresses the problem of interpreting user intentions through two primary methods. The first method (true NLU) uses a combination of keyword analysis, semantic processing, discourse processing, and context analysis to determine the meaning and intent of speech [15]. This second method (phrase mapping) is less complex and requires fewer computational resources than true NLU, making it suitable for real-time intent determination.

With high-level speech commands as an input, control systems should translate wearer intents into low-level controllers, such as position, force, or impedance controllers, to synchronize gait planning in a smooth and time-continuous manner. Finite state machines (FSMs) address this motion planning problem by acting as a central planner for transitions between exoskeleton states, such as standing and walking [16].

One bioinspired strategy for exoskeleton control is the central pattern generator (CPG). The CPG consists of connected nodes that can generate rhythmic patterns without receiving rhythmic inputs, facilitating joint motion synchronization necessary to replicate rhythmic motor behaviors such as bipedal

locomotion [17], [18]. CPGs typically include parameters that allow for modulation of locomotion, which provides additional control and flexibility for the wearer [19], [20]. The ability of CPGs to generate time-continuous rhythmic motions makes them an appropriate candidate for shaping the trajectories of lower limb exoskeletons, as shown in [21], [22], [23], [24], [25].

In this study, a speech-based locomotion planning strategy combined social HRI, FSM, and CPG for the intelligent motion planning of a lower limb exoskeleton for bipedal locomotion. As opposed to a button-based interface, voice input was the primary mode of determining user intent. These high-level commands were processed by an FSM to ensure safe and natural state transitions before being executed in low-level position controllers. The major contributions of the proposed control scheme can be summarized as follows:

- Integration of SR and a lower limb exoskeleton to create a system that allows a user’s hands to be free to use mobility aids while still controlling the exoskeleton.
- A novel set of CPG dynamics was proposed to synchronize time-continuous transitions between exoskeleton locomotion states (e.g., sit, stand, walk) in response to discrete user inputs. Speech inputs were processed through an FSM alongside joint angles and velocities to streamline state transitions (e.g., speed up, slow down). Although voice-activated robotic systems have previously been investigated [26], [27], previous CPG dynamics have not incorporated speech-based inputs for lower limb exoskeletons [17].

II. SPEECH PROCESSING SYSTEM

The proposed speech processing system was a customized version of the pipeline (Fig. 1) developed by Reverb Robotics [28], consisting of a denoising function to increase the signal-to-noise ratio (SNR), a speech perception module to process raw speech data, and a natural language understanding model to determine speech intent. The system is compatible with a Raspberry Pi. Once an intent was determined, the Pi sent the command over a network to the central exoskeleton motion planner as shown in Fig. 1.

A. Denoising Pipeline

The denoising function processed raw user inputs before passing the denoised signal to the speech perception function. Denoising was achieved by detecting voice activity, analyzing its spectral signature to estimate noise, and subtracting its spectral signature of the noise from the input audio to produce the denoised audio (Fig. 2). We used RNNoise, a deep recurrent neural network, to denoise the input SR audio [29].

B. Speech Perception Network

The speech processing system used the Vosk speech perception system [30]. The Vosk implementation combines the time delay neural network factorization model and a multi-stream convolutional neural network. We used a pre-trained set

TABLE I
CPG PARAMETERS, INITIAL CONDITIONS, AND CONSTANTS FOR THE PROPOSED MOTION PLANNING SCHEME. l AND r REFER TO THE LEFT AND RIGHT SIDE OF THE EXOSKELETON, RESPECTIVELY.

Parameters	Initial conditions	Constants
$v_{ij} = 0.1$	$A_0(k) = 1$	$c_r = 2.5$
$\phi_{ll} = 0 \text{ rad}$	$\Omega_0(k) = \pi/2 \text{ rad}$	$c_\theta = 2 \text{ rad}$
$\phi_{rr} = 0 \text{ rad}$	$\theta_l(0) = (2 + \pi) \text{ rad}$	$\beta_\omega = 10\pi$
$\phi_{lr} = \pi \text{ rad}$	$\theta_r(0) = 2 \text{ rad}$	$\beta_r = 10\pi$
		$T = 2 \text{ s}$

of open-source weights called ‘vosk-model-small-en’ provided by Alpha Cephei [31].

C. Natural Language Understanding

The SR system used Snips NLU [32] to create a list of intents along with phrases that trigger those intents. The intent parser used a probabilistic engine that focuses on keyword analysis, enabling the NLU engine to associate perceived speech with the proper intent. Once an intent was extracted, the program determined whether the speech was directed at the exoskeleton by checking if the command contained the word ‘robot’. The specific key words or phrases to trigger a state transition were ‘robot {keep moving, don’t change, maintain speed, stand up, stand, sit down, sit, stop moving, stop, walk forward, walk, move forward, move, forward, slow down, slow, speed up, go faster, faster}’, where one of the key words or phrases in the set are used (Fig. 3).

III. MOTION PLANNING AND CONTROL

A. Overview of the Exoskeleton Motion Planning System

The proposed exoskeleton locomotion motion planning strategy combined a speech processing system, FSM, and CPG to ensure safe time-continuous transitions between exoskeleton states (Fig. 1).

B. Synchronization of Joint Trajectories

The proposed CPG dynamics for the phase $\theta_i(t)$ and amplitude $r(t)$ of the i th exoskeleton joint are governed by the following equations based on the Kuramoto model for the synchronization of coupled oscillators [33].

$$\begin{aligned} \dot{\theta}_i(t) &= \omega(t) + \sum_{j=1}^N v_{ij} \sin(\theta_i(t) - \theta_j(t) - \phi_{ij}) \\ \ddot{\omega}(t) &= \lambda(t)\beta_\omega \left(\frac{\beta_\omega}{4} (\Omega_n(k) - \omega(t)) - \dot{\omega}(t) \right) \\ \ddot{r}(t) &= \lambda(t)\beta_r \left(\frac{\beta_r}{4} (A_n(k) - r(t)) - \dot{r}(t) \right) \end{aligned} \quad (1)$$

where N is the number of joints, v_{ij} is the coupling strength, ϕ_{ij} is the phase offset, β_ω and β_r are fixed constants, $\Omega_n(k)$ and $A_n(k)$ are user-adjustable constants which modulate the frequency and amplitude of the system, respectively, and $\lambda(t)$ is a user-triggered ramping system which multiplies the CPG

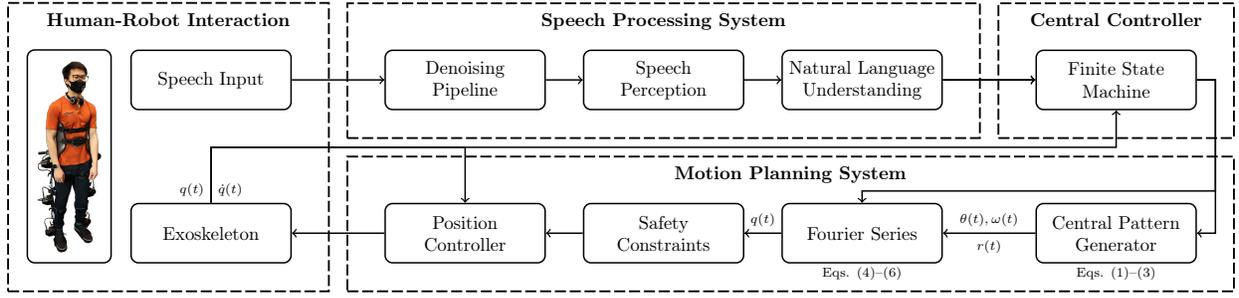


Fig. 1. Proposed strategy for speech-based planning of lower limb exoskeletons.

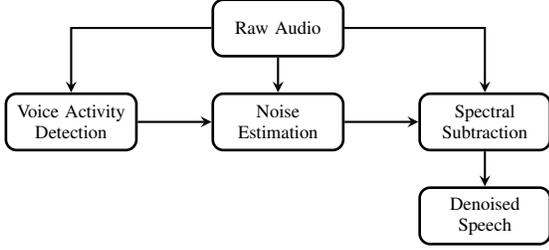


Fig. 2. Major functions for denoising voice activity.

signal by a linear time-dependent gain. In particular, $\Omega_n(k)$ and $A_n(k)$ update in response to user inputs, k (Table I).

$$\Omega_n(k) = \begin{cases} \Omega_{n-1} + c_\theta, & k = \text{speed up} \\ \Omega_{n-1} - c_\theta, & k = \text{slow down} \\ \Omega_{n-1}, & k = \text{otherwise} \end{cases} \quad (2)$$

$$A_n(k) = \begin{cases} A_{n-1} + c_r, & k = \text{speed up} \\ A_{n-1} - c_r, & k = \text{slow down} \\ A_{n-1}, & k = \text{otherwise} \end{cases}$$

Here, Ω_n and A_n are the updated speed constants, Ω_{n-1} and A_{n-1} are the current speed constants, and c_θ and c_r are constants that adjust the frequency and amplitude of the dynamics in (1).

The ramping system is defined as follows:

$$\lambda(t) = \begin{cases} t/T, & \text{stand-to-walk} \\ 1 - t/T, & \text{walk-to-stop} \\ 1, & \text{walking} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where the stand-to-walk and walk-to-stop conditions are triggered by user commands and held for a period T (Table I). Coupling between all joints is maintained by the same principal frequency of $\omega(t)$ and amplitude $r(t)$ to synchronize locomotion trajectories.

The desired walking trajectory $q_{w_i}(t)$ for the joint i of the exoskeleton is defined as

$$q_{w_i}(t) = r(t) \left(a_{0_i} + \sum_{k=1}^{N_i} (a_{k_i} \cos k\theta_i(t) + b_{k_i} \sin k\theta_i(t)) \right) \quad (4)$$

where a_{k_i} and b_{k_i} are the coefficients of Fourier series with N_i terms (Table II). The amplitude and phase of these oscillatory

motions are updated in real-time by $\theta_i(t)$ and $r(t)$. Trajectories for the ankle, knee, and hip during walking were obtained from Subject 6 in the experiments of [34].

The sit-to-stand and stand-to-sit trajectories for the ankle, knee, and hip are obtained from the experiments of [35]. The mean trajectory values of all subjects for each joint were used to compute the desired sit-to-stand and stand-to-sit trajectory $q_{s_i}(t)$ for the joint i of the exoskeleton.

$$q_{s_i}(t) = a_{0_i} + \sum_{k=1}^{N_i} (a_{k_i} \cos k\omega_{s_i} t + b_{k_i} \sin k\omega_{s_i} t) \quad (5)$$

where ω_{s_i} is the angular frequency of the trajectory, and a_{k_i} and b_{k_i} are the coefficients of the Fourier series with N_i terms (Table II). Equation (5) was set to satisfy the time-dependent boundary conditions $\dot{q}_{s_i}(t_0) = 0$ and $\dot{q}_{s_i}(T) = 0$, where t_0 is the initial time and T is the time after one sit-to-stand or stand-to-sit period (Table I). The output of (5) was transformed into the coordinate system of the exoskeleton via the linear function

$$q_{t,s_i}(t) = -q_{s_i}(t) + \max(q_{s_i}(t)) \quad (6)$$

where $q_{t,s_i}(t)$ is the transformed angle. This function ensures that the endpoint for the sit-to-stand trajectory for the ankle, knee, and hip terminates at 0° . The stand-to-sit trajectory is implemented as the reverse of the sit-to-stand trajectory.

C. Safety Considerations

The finite state machine was designed to ensure safe transitions between exoskeleton states. The states and intents for triggering transition can be seen in figure 3, and the initial state of the exoskeleton is either sitting or standing, depending on the initial position of the user. Position controller constraints include limitations on the maximum torque and velocity and maximum and minimum joint angles.

IV. EXPERIMENTAL EVALUATIONS

A. Speech Processing Experiments

Experimental assessments were performed to evaluate the efficacy of the SR system with ethics permission from the University of Lethbridge Human Subjects Review Board, number 2013-037. In the experiment, participants spoke a series of predefined commands consistent with normal use of the exoskeleton in a quiet environment. After each command, the output of the SR system and NLU engine was recorded.

TABLE II
COEFFICIENTS OF THE FOURIER SERIES FOR THE HIP AND KNEE FROM SITTING, STANDING, AND WALKING.*

State	Joint	a_0	a_1	a_2	a_3	a_4	a_5	a_6	b_1	b_2	b_3	b_4	b_5	b_6
Sitting/ Standing	Hip	105.40	-1.52	1.29	0.42	0.36	-0.04	-0.12	-3.86	2.92	0.24	-0.49	-0.28	0.01
	Knee	140.20	-38.30	-1.75	0.55	1.32	0.14	n/a	-29.22	8.39	4.14	0.15	-0.39	n/a
	Ankle	49.59	22.01	-1.31	0.28	-0.07	0.15	n/a	35.11	-11.48	-4.05	-0.16	0.40	n/a
Walk	Hip	40.69	23.22	-4.49	0.40	0.70	1.08	-0.27	-8.65	3.34	1.39	0.80	0.34	0.07
	Knee	25.70	-3.83	-8.54	1.91	1.09	2.05	-0.31	-19.28	17.93	3.77	1.50	0.58	-0.90
	Ankle	-0.99	5.29	-8.58	-0.48	1.69	-0.04	1.30	3.61	-5.95	5.92	-2.12	1.02	-0.72

* $R_{adj}^2 \geq 0.99$ for each trajectory.

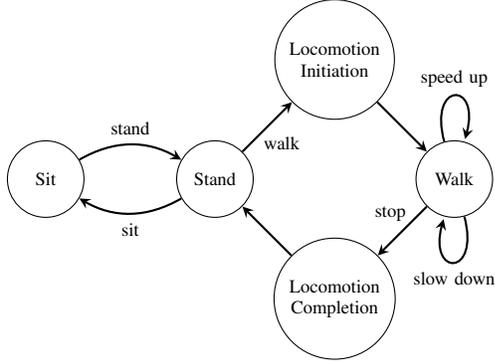


Fig. 3. Proposed FSM to plan the transitions between exoskeleton states.

Then, the participant put on a face mask and repeated the experiment. SR performance was measured with two metrics, word error rate (WER) and intent error rate (IER). WER is determined by the ratio of insertions (I), substitutions (S), and deletions (D) required to transform a response into the target to the number of words in the target N :

$$\text{WER} = \frac{I + S + D}{N} \times 100\% \quad (7)$$

IER is a measure of whether the correct intent was derived from the user's speech:

$$\text{IER} = \frac{1}{N} \left(\sum_{\text{all trials}} \begin{cases} 0, & \text{correct output} \\ 1, & \text{incorrect output} \end{cases} \right) \times 100\% \quad (8)$$

After WER and IER were calculated for individual trials, the results were aggregated to get a cumulative average for both metrics.

Experiments were performed by 4 male and 3 female participants. Participants were taken from a population of undergraduate and graduate students at the University of Lethbridge aged between 19 and 31. WER and IER scores from each participant were aggregated and computed (Fig. 4).

B. Locomotion Tasks With and Without Speech Commands

The proposed motion planning strategy was assessed experimentally to provide proof-of-concept evidence for the effectiveness of speech-based locomotion planning with the Exo-H3 lower limb exoskeleton from Technaid. The proposed motion planning scheme was implemented in real-time using MATLAB Simulink, which received sensory data and controlled motors at a sampling rate of 100 Hz via a CAN

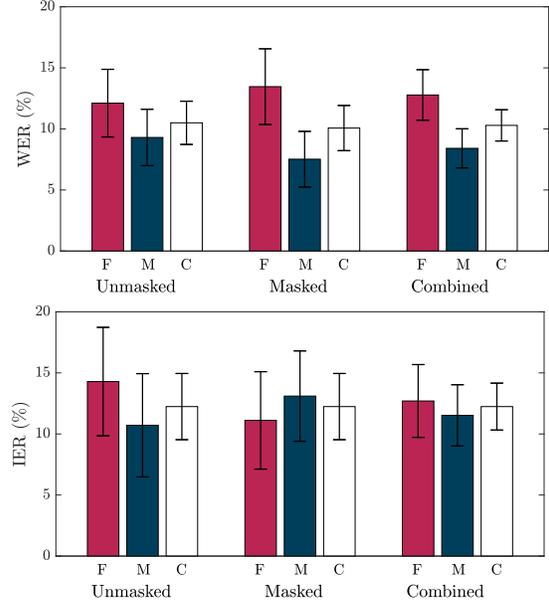


Fig. 4. Results for the human SR trials for WER (top) and IER (bottom). F=Female, M=Male, C=Combined. There were no significant differences between or within groups.



Fig. 5. Exo-H3 lower limb exoskeleton worn by two able-bodied users for walking. The participants use a wireless headset to command the exoskeleton. (a) User 1 (21-year-old) and (b) User 2 (33-year-old).

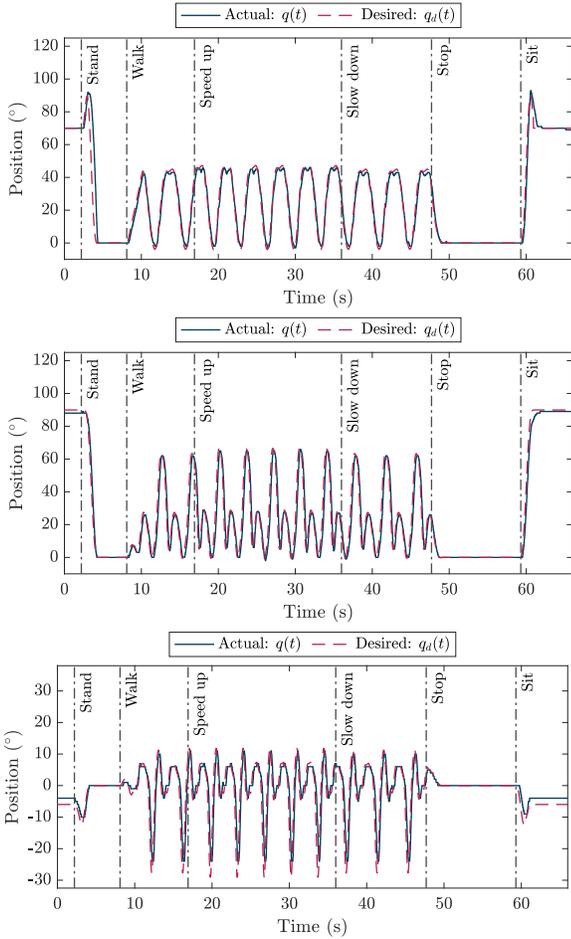


Fig. 6. Actual and desired trajectories for the left hip (top), knee (middle), and ankle (bottom) for A to B locomotion task for user 1.

interface (Vector VN1610) with 2 channels. The SR system was run on a Raspberry Pi 3, which received speech input from the microphone on a Logitech Wireless Headset H600 at a sampling rate of 16 000 Hz, and the processed signals were transmitted to the laptop via the user datagram protocol. The major computations in this strategy involve time integration of the CPG dynamics in equation (1) and calculating each joint angle for the exoskeleton with Fourier series in equations (4), (5), and (6). Preliminary tests were performed in a trial-and-error manner to obtain reasonable parameters for speed modulation (c_θ and c_r in equation 2) and initial values and parameters of the proposed CPG system (Table I). These final CPG parameters were used in the described experiments. Here, two able-bodied users wore the exoskeleton using a walker for postural stability (Fig. 5). Various angles for a forward-learning posture were experimentally trialed (0° to 30°), and 15° resulted in a subjective optimal angle.

The first experiment involved walking in a straight line for approximately 12 m. The participants were initially sitting, stood and walked, then sat. Between the start and end points, the user could choose to speed up or slow down ad-lib. A 21-year-old participant, user 1, and a 33-year-old participant, user 2, completed 9 and 12 trials, respectively. The time to walk

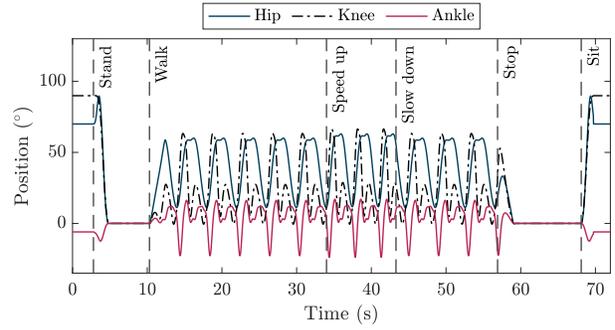


Fig. 7. Desired trajectories for the left hip, knee, and ankle for A to B locomotion task for user 2.

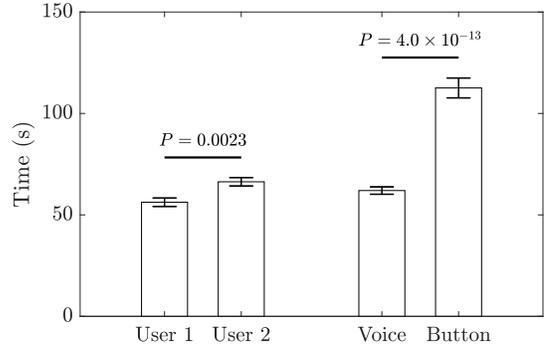


Fig. 8. Time taken by users 1 and 2 to complete a locomotion task using voice control ($n_1 = 9$, $n_2 = 12$). Voice is the weighted mean time of users 1 and 2 to complete the task with voice control. Button is the mean time of user 1 to complete the task with a button-based interface ($n = 10$).

in these trials for user 1 was 56 ± 2 s and that for user 2 was 66 ± 2 s ($P = 0.0023$ for a two-tailed t-test assuming equal variances, Fig. 8). The significant difference in time between users 1 and 2 to complete this experiment can be attributed to user preferences in walking speeds. The trajectory for walking from point A to point B for one representative trial of user 1 is plotted in Fig. 6 and that for user 2 is plotted in Fig. 7.

The same experiment was performed using a button-based smartphone app from Technaid with user 1 as the participant for 10 trials, and the time taken to walk from A to B was 113 ± 5 s. In comparison with the voice command time of 62 ± 2 s, a two-tailed t-test assuming equal variances yielded $P = 4.0 \times 10^{-13}$ (Fig. 8). This significant difference can be attributed to the additional time the user needed to stop the exoskeleton, remove their hands from their walker, search for the appropriate command on the exoskeleton remote control, and press that button.

V. CONCLUSIONS

In this study, we developed a speech-based locomotion planning strategy to provide safe and convenient motion planning of a lower limb exoskeleton. This method combined a speech processing system, FSM, and novel CPG to plan the exoskeleton movements based on user intents.

The proposed speech processing system showed WER and IER values of 10.29% and 12.24%, respectively, tested with participants aged 19-31 under various conditions. Despite

being slightly higher than desktop models cited in [36], its performance suggests potential for speech-based HRI applications, especially on limited hardware.

In user studies, the exoskeleton's voice control outperformed button control in locomotion tasks, being 54% faster. Voice commands also freed users' hands, improving postural stability. Thus, speech-based planning increased convenience and safety, making it a potential choice for the control of lower limb exoskeletons.

REFERENCES

- [1] A. Rodríguez-Fernández, J. Lobo-Prat, and J. M. Font-Llagunes, "Systematic review on wearable lower-limb exoskeletons for gait training in neuromuscular impairments," *Journal of NeuroEngineering and Rehabilitation*, vol. 18, 2021.
- [2] G. Zeilig, H. Weingarden, M. Zwecker, I. Dudkiewicz, A. Bloch, and A. Esquenazi, "Safety and tolerance of the rewalk™ exoskeleton suit for ambulation by people with complete spinal cord injury: A pilot study," *The Journal of Spinal Cord Medicine*, vol. 35, no. 2, pp. 96–101, 2012.
- [3] S. A. Murray, R. J. Farris, M. Golfarb, C. Hartigan, C. Kandilakis, and D. Truex, "Fes coupled with a powered exoskeleton for cooperative muscle contribution in persons with paraplegia," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 2788–2792.
- [4] O. Jansen, D. Grasmuecke, R. C. Meindl, M. Tegenthoff, P. Schwenkreis, M. Sczesny-Kaiser, M. Wessling, T. A. Schildhauer, C. Fisahn, and M. Aach, "Hybrid assistive limb exoskeleton hal in the rehabilitation of chronic spinal cord injury: Proof of concept; the results in 21 patients," *World Neurosurgery*, vol. 110, pp. e73–e78, 2018.
- [5] K. A. Inkol and J. McPhee, "Assessing control of fixed-support balance recovery in wearable lower-limb exoskeletons using multibody dynamic modelling," in *2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechanics (BioRob)*, 2020, pp. 54–60.
- [6] T. Cheng and M. Sharifi, "Development of an assistive ankle-foot exoskeleton with sensorized silicone-based insole," *ASME Journal of Engineering and Science in Medical Diagnostics and Therapy*, vol. 7, no. 2, p. 021010 (7 pages), 2024.
- [7] J. E. Patricio, M. Sharifi, D. Shrestha, S. H. S. Thu, and E. Kwan, "Design and development of a lightweight, high-torque, and cost-effective hip exoskeleton," in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, 2023, pp. 367–372.
- [8] D. M. Rivera and M. Sharifi, "Design and fabrication of a lightweight and wearable semirigid robotic knee chain exoskeleton," *ASME Journal of Engineering and Science in Medical Diagnostics and Therapy*, vol. 7, no. 2, p. 021007 (8 pages), 2024.
- [9] V. Villani, F. Pini, F. Leali, and C. Secchi, "Survey on human-robot collaboration in industrial settings: Safety, intuitive interfaces and applications," *Mechatronics*, vol. 55, pp. 248–266, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957415818300321>
- [10] T. B. Sheridan, "Human-robot interaction: Status and challenges," *Human Factors*, vol. 58, no. 4, pp. 525–532, 2016.
- [11] J. Kennedy, S. Lemaignan, C. Montassier, P. Lavalade, B. Irfan, F. Papadopoulous, E. Senft, and T. Belpaeme, "Child speech recognition in human-robot interaction: Evaluations and recommendations," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 82–90.
- [12] X. Huang, J. Baker, and R. Reddy, "A historical perspective of speech recognition," *Commun. ACM*, vol. 57, no. 1, p. 94–103, jan 2014. [Online]. Available: <https://doi.org/10.1145/2500887>
- [13] T. F. Abidin, A. Misbullah, R. Ferdhiana, M. Z. Aksana, and L. Farsiah, "Deep neural network for automatic speech recognition from indonesian audio using several lexicon types," in *2020 International Conference on Electrical Engineering and Informatics (ICELTICs)*, 2020, pp. 1–5.
- [14] N. Moritz, J. Schröder, S. Goetze, J. Anemüller, and B. Kollmeier, "Acoustic scene classification using time-delay neural networks and amplitude modulation filter bank features," *complexity*, vol. 12, p. 13, 2016.
- [15] M. Bates, "Models of natural language understanding," *Proceedings of the National Academy of Sciences*, vol. 92, no. 22, pp. 9977–9982, 1995.
- [16] K. A. Strausser and H. Kazerooni, "The development and testing of a human machine interface for a mobile medical exoskeleton," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011, pp. 4911–4916.
- [17] A. J. Ijspeert, "Central pattern generators for locomotion control in animals and robots: A review," *Neural Networks*, vol. 21, no. 4, pp. 642–653, 2008, robotics and Neuroscience.
- [18] J. Arcos-Legarda, J. Cortes-Romero, and A. Tovar, "Robust compound control of dynamic bipedal robots," *Mechatronics*, vol. 59, pp. 154–167, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S095741581930039X>
- [19] M. Sharifi, J. K. Mehr, V. K. Mushahwar, and M. Tavakoli, "Autonomous locomotion trajectory shaping and nonlinear control for lower-limb exoskeletons," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 2, pp. 645–655, 2022.
- [20] S. O. Schrader, Y. Nager, A. R. Wu, R. Gassert, and A. Ijspeert, "Bio-inspired control of joint torque and knee stiffness in a robotic lower limb exoskeleton using a central pattern generator," in *2017 International Conference on Rehabilitation Robotics (ICORR)*, 2017, pp. 1387–1394.
- [21] M. Sharifi, J. K. Mehr, V. K. Mushahwar, and M. Tavakoli, "Adaptive cpg-based gait planning with learning-based torque estimation and control for exoskeletons," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 8261–8268, 2021.
- [22] K. Gui, H. Liu, and D. Zhang, "A generalized framework to achieve coordinated admittance control for multi-joint lower limb robotic exoskeleton," in *2017 International Conference on Rehabilitation Robotics (ICORR)*, 2017, pp. 228–233.
- [23] D. Zhang, Y. Ren, K. Gui, J. Jia, and W. Xu, "Cooperative control for a hybrid rehabilitation system combining functional electrical stimulation and robotic exoskeleton," *Frontiers in Neuroscience*, vol. 11, p. 725, 2017.
- [24] J. K. Mehr, M. Sharifi, V. K. Mushahwar, and M. Tavakoli, "Intelligent locomotion planning with enhanced postural stability for lower-limb exoskeletons," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7588–7595, 2021.
- [25] X. Zhang and M. Hashimoto, "Synchronization-based trajectory generation method for a robotic suit using neural oscillators for hip joint support in walking," *Mechatronics*, vol. 22, no. 1, pp. 33–44, 2012. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957415811001723>
- [26] Y. Guo, W. Xu, S. Pradhan, C. Bravo, and P. Ben-Tzvi, "Personalized voice activated grasping system for a robotic exoskeleton glove," *Mechatronics*, vol. 83, p. 102745, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957415822000058>
- [27] X. Wang, P. Tran, S. Callahan, S. Wolf, and J. Desai, "Towards the development of a voice-controlled exoskeleton system for restoring hand function," 04 2019, pp. 1–7.
- [28] L. S. Grasse and M. S. Tata, "An end-to-end platform for human-robot speech interaction," *Sound in HRI Workshop. 16th Annual Conference for Basic and Applied Human-Robot Interaction.*, Mar 2021.
- [29] J.-M. Valin, "A hybrid dsp/deep learning approach to real-time full-band speech enhancement," 2018.
- [30] K. J. Han, J. Pan, V. K. N. Tadala, T. Ma, and D. Povey, "Multistream cnn for robust acoustic modeling," 2021.
- [31] "Vosk models." [Online]. Available: <https://alphacephei.com/vosk/models>
- [32] A. Coucke, A. Saade, A. Ball, T. Bluche, A. Caulier, D. Leroy, C. Doumouro, T. Gisselbrecht, F. Caltagirone, T. Lavril, M. Primet, and J. Dureau, "Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces," *CoRR*, vol. abs/1805.10190, 2018.
- [33] F. A. Rodrigues, T. K. D. Peron, P. Ji, and J. Kurths, "The kuramoto model in complex networks," *Physics Reports*, vol. 610, pp. 1–98, 2016, the Kuramoto model in complex networks.
- [34] T. Lencioni, I. Carpinella, M. Rabuffetti, A. Marzegan, and M. Ferrarin, "Human kinematic, kinetic and emg data during different walking and stair ascending and descending tasks," *Scientific data*, vol. 6, no. 1, Dec. 2019.
- [35] S. Nuzik, R. Lamb, A. VanSant, and S. Hirt, "Sit-to-Stand Movement Pattern: A Kinematic Study," *Physical Therapy*, vol. 66, no. 11, pp. 1708–1713, 11 1986.
- [36] X. Liu, P. He, W. Chen, and J. Gao, "Multi-task deep neural networks for natural language understanding," 2019. [Online]. Available: <https://arxiv.org/abs/1901.11504>