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 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 \( r_i(t) \) of the \( i \)th exoskeleton joint are governed by the following equations based on the Kuramoto model for the synchronization of coupled oscillators [33].

\[
\dot{\theta}_i(t) = \omega(t) + \sum_{j=1}^{N} v_{ij} \sin(\theta_i(t) - \theta_j(t) - \phi_{ij}) \\
\dot{\omega}(t) = \lambda(t) \beta_\omega \left( \frac{3}{4} (\Omega_n(k) - \omega(t)) - \dot{\omega}(t) \right) \\
\dot{r}(t) = \lambda(t) \beta_r \left( \frac{3}{4} (A_n(k) - r(t)) - \dot{r}(t) \right)
\]

where \( N \) is the number of joints, \( v_{ij} \) is the coupling strength, \( \phi_{ij} \) is the phase offset, \( \beta_\omega \) and \( \beta_r \) 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 dynamics.
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).

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

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

(2)

Here, $\Omega_n$ and $A_n$ are the updated speed constants, $\Omega_{n-1}$ and $A_{n-1}$ are the current speed constants, and $c_\theta$ and $c_r$ are constants that adjust the frequency and amplitude of the dynamics in (1).

The ramping system is defined as follows:

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

(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

$$q_{w_i}(t) = r(t) \left( a_0 + \sum_{k=1}^{N_i} (a_k \cos k\theta_i(t) + b_k \sin k\theta_i(t)) \right)$$

(4)

where $a_k$ and $b_k$ 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.

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

(5)

where $\omega_{s_i}$ is the angular frequency of the trajectory, and $a_{k_i}$ and $b_{k_i}$ are the coefficients of the Fourier series with $N_i$ (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

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

(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 $0^\circ$. 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.
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.
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 ($c_\theta$ 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 ($0^\circ$ to $30^\circ$), and $15^\circ$ 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 in these trials for user 1 was $56 \pm 2\, s$ and that for user 2 was $66 \pm 2\, s$ ($P = 0.0023$ 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 $113 \pm 5\, s$. In comparison with the voice command time of $62 \pm 2\, s$, 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