

راهکار کنترل مقاوم مبتنی بر یادگیری تقویتی به منظور توانبخشی حرکتی بازوی دست

زهرا حسن‌زاده بناییدی^۱، حمیدرضا کبروی^۲، سعید طوسی‌زاده^۳، رضا بوستانی^۴

دانشجو کارشناسی ارشد مهندسی پزشکی، گروه مهندسی پزشکی، دانشگاه آزاد اسلامی، واحد مشهد، hasanzadeh_67511@yahoo.com

استادیار، دانشکده فنی مهندسی، گروه مهندسی پزشکی، دانشگاه آزاد اسلامی، واحد مشهد، hkobravi@mshdiau.ac.ir

استادیار، دانشکده فنی مهندسی، گروه مهندسی برق، دانشگاه آزاد اسلامی، واحد مشهد، saeedtoosizadeh@yahoo.com

استادیار، گروه مغز و اعصاب، دانشگاه علوم پزشکی مشهد، BoostaniR@mums.ac.ir

(تاریخ دریافت مقاله ۱۳۹۲/۵/۲۳، تاریخ پذیرش مقاله ۱۳۹۲/۷/۲۸)

چکیده: توانبخشی حرکتی از مباحث مورد توجه محققان است. در این تحقیق، یک راهکار کنترلی به منظور کنترل حرکت مدلی از بازوی دست با سه مفصل ارائه شده است. در مدل مورد استفاده، اثر فعالسازی همزمان عضلات آگونیست و آنتاگونیست مفصل مج دست با استفاده از مکانیزم فریزسازی لحظه شده است. با الهام از عملکرد سیستم اعصاب مرکزی در کسب مهارت‌های حرکتی، راهکار کنترلی ارائه شده مبتنی بر یکی از الگوریتم‌های یادگیری تقویتی توسعه داده شده است. ابتدا کارایی کنترل کننده کلاسیک مبتنی بر یادگیری تقویتی به منظور کنترل حرکت دست به سمت یک هدف مشخص، مورد ارزیابی قرار گرفت. نتایج نشان دادند که چنین کنترل کننده‌ای نمی‌تواند حتی بازای پس چندین اپیزود پایداری سیستم را حفظ نماید. در مرحله بعد به منظور بهبود وضعیت پایداری و همچنین کارایی کنترل کننده، ترکیبی از کنترل کننده تنسی‌انتگرالی-مشتقی (PID) و کنترل کننده مبتنی بر یادگیری تقویتی ارائه و مورد ارزیابی قرار گرفت. کنترل کننده PID در کنار یادگیری تقویتی کارایی خوبی داشته است، اما به لحاظ تئوری تضمینی برای اثبات پایداری وجود ندارد. لذا به منظور بهبود شرایط پایداری سیستم کنترل، از ترکیب کنترل کننده HTC و کنترل کننده PID همراه با کنترل کننده مبتنی بر راهکار یادگیری تقویتی استفاده شده است. به این ترتیب پایداری سیستم کنترل حلقه بسته مبتنی تئوری لیپاتوف تضمین شده است. نتایج نشان دادند که کنترل کننده ترکیبی با استفاده از HTC در برابر اختشاش خارجی و تغییرات تصادفی پارامترهای سیستم، کارایی بسیار خوبی داشته است. پس از شیوه‌سازی-های کامپیوتری، به منظور ارزیابی کامپیوتری راهکار کنترلی پیشنهادی، آزمایش‌های انسانی طراحی و انجام شدند. در آزمایش‌های انسانی الگوی فعالیت عضلات مفاصل دست، طی عملکرد حرکت دست با استفاده از سیگنال‌های الکترومایوگرام عضلات، استخراج شدند. نتایج نشان می‌دهند که همزمانی قابل ملاحظه‌ای بین بازه زمانی مربوط به فعالیت و عدم فعالیت تغییرات سیگنال خروجی کنترل کننده در مطالعات شبیه‌سازی با الگوی فعالیت عضلات دست فرد سالم وجود دارد.

کلید واژه: کنترل کننده تنسی‌انتگرالی-مشتقی، کنترل کننده مقاوم (HTC)، مدل بازوی سه مفصل، مکانیزم فریزسازی، یادگیری تقویتی

A Robust Control Strategy Based on Reinforcement Learning Approach to Rehabilitate the Arm Movement

Z. Hasanzadeh Binabidi, Hamid Reza Kobravi, Saeed Toosizadeh, Reza Boostani

In this research, a control strategy has been presented to movement control of a three link model of human's arm. The freezing mechanism has been used to consider the role of antagonistic coactivation of wrist muscles in the used three link model. Inspired by motor learning process of central nervous system, the presented control strategy has been designed based on the reinforcement learning algorithm. At first, the performance of a control methodology based on reinforcement learning was evaluated. The results show the instability of control system even after numbers of leaning episode. Then, a combination of a proportional derivative integral (PID) controller and a reinforcement learning based controller were utilized to improve the stability conditions and performance of controller. Despite the good performance, there is no guarantee for stability of control system. So, to satisfy the stability conditions, a robust controller called HTC was added to the combination of a PID controller and a reinforcement learning based controller. According to the simulation results, the combinational controller accompany by HTC had good performance even in presence of external disturbance, measurement noise and random changes of model parameters. For more assessments, the muscle activation profile of involved muscles during the arm movement of an intact subject was compared with control signals obtained through the simulation studies. The results show an interesting timing synchronization between the activation and deactivation timing of control signals and muscle activation profiles.

Keyword: proportional derivative integral controller, robust controller (HTC), three link model of human's arm, freezing mechanism, reinforcement learning

مدل دو سطحی، شامل مدل دینامیکی سطح بالا و مدل دینامیکی سطح پایین هستند که با یکدیگر هماهنگی دارند. در تحقیق دیگری جانگ یو و همکارانش، تئوری کنترل ثابت بهینه پیوسته در زمان را برای مدل‌سازی سیستم‌های کنترل حرکتی انسان در حضور نویز وابسته به سیگنال، ارائه داده‌اند [۲]. مدل کنترلی آنها با برخی از مشاهدات مهم مانند عدم تقارن پروفایل سرعت سازگار است. جاجودنیک و همکاران سه کنترلر پس‌خوردی تناسبی-مشتقی (PD) برای تحریک شش عضله بازو، با استفاده از دو سنسور زاویه مفصل ارائه داده‌اند [۸]. استفاده از راهکارهای کنترل ترکیبی مانندتر کیپ کنترل پس‌خوردی و جلوسوی، یادگیری تقویتی و شبکه‌های عصبی مصنوعی نیز مورد ارزیابی قرار گرفته‌اند [۹]. اما با توجه به اینکه در فرآیند حرکت دست به سمت هدف، اولاً خط‌سیر مطلوبی برای تغییرات موقعیت دست قابل طراحی نیست و ثانیاً موقعیت نقاط ابتداء و انتهای نیز ثابت نیستند، لذا استفاده از روش‌های کنترل مبتنی بر یادگیری تقویتی می‌تواند موثرتر باشد [۱۷]-[۱۰]. یادگیری تقویتی به معنای تصحیح عملکرد یک سیستم براساس پاداش و تنبیه‌هایی است که با توجه به شناخت از محیط و نتایج تعامل با محیط، طراحی شده‌اند. پژوهشگران دریافت‌های اند که نرخ ارسال الکتریکی نورون‌های پخش کننده دوپامین در ناحیه تگمتومن شکمی را می‌توان به تابع خطاگری تمثیل کرد [۱۹]، که در فرآیند حرکت دست به سمت هدف، این تابع در حقیقت وابسته به فاصله بین موقعیت فعلی و نهایی است. این امر موجب افزایش غایلیت مغزی در ناحیه گانگلیون بازال^۴ می‌گردد، که این پخش از ساختار مغز در کنترل حرکات عضله و یادگیری فرد دخالت دارد [۱۹]. با الهام از عملکرد سیستم اعصاب مرکزی در کسب مهارت‌های حرکتی، راهکار کنترلی مبتنی بر یکی از الگوریتم‌های یادگیری تقویتی توسعه داده شده است. به دلیل وجود عدم قطعیت، دینامیک‌های مدل نشده و عدم وجود مدل ریاضی مناسب در سیستمهای اسکلتی-عضلانی، بکارگیری روش‌های یادگیری تقویتی در کنار کنترل کننده‌هایی مقاوم می‌تواند، راهکاری موثر باشد. بر این اساس هدف اصلی این تحقیق، ارائه راهکار کنترلی جدیدی بر مبنای ترکیب یک کنترلگر با الگوریتم‌های یادگیری تقویتی می‌باشد. به منظور بهبود وضعیت پایداری و همچنین کارایی کنترل-کننده، ترکیبی از یک کنترل کننده پس‌خوردی و کنترل کننده مبتنی بر یادگیری تقویتی (RLPID) مورد استفاده قرار گرفته است. سپس به منظور تضمین کاملتر پایداری، ساختار کنترلی شامل ترکیب کنترل کننده مقاوم و کنترل کننده پس‌خوردی همراه با کنترل کننده مبتنی بر راهکار یادگیری تقویتی (RLPIDHTC) استفاده شده است. به این ترتیب پایداری سیستم کنترل حلقه بسته با استفاده از تئوری لیپانوف تضمین می‌شود. نتایج مطالعات مبتنی بر شیوه‌سازی‌های کامپیوترا نشان می‌دهند که راهکار کنترلی RLPIDHTC، دارای بهترین کارایی بوده است. پس از ارزیابی بوسیله شیوه‌سازی کامپیوترا، به منظور ارزیابی کاملتر راهکار کنترلی پیشنهادی، آزمایش‌های انسانی طراحی و انجام شده‌اند. بطوریکه

۱- مقدمه

آسیب نخاعی^۱ سطح بالا (سطح گردنی C1-C4)، می‌تواند موجب از دست دادن اکثر یا همه عملکردهای عضلانی ارادی در زیر گردن می‌باشد. تحریک الکتریکی عملکردی^۲ می‌تواند به منظور ایجاد عملکردهای حرکتی در این افراد موثر باشد. در این راهکار با تحریک الکتریکی اعصاب حرکتی عضلات فلج و تنظیم شدت انقباض آنها، گشاوارهای مکانیکی حول مفاصل، کنترل و به طبع آن حرکت ایجاد می‌شود. بدلیل دینامیک متغیر، اغتشاشات، رفتار متغیر با زمان و عدم قطعیت‌های موجود در سیستمهای اسکلتی-عضلانی، توسعه سیستمهای مبتنی بر FES نیازمند استفاده از روش‌های کنترلی هستند. یکی از زمینه‌های کاربردی تحریک الکتریکی عملکردی در ایجاد و کنترل حرکت بازو دست است. حرکت بازو یکی از رایج‌ترین فعالیت‌ها در زندگی روزمره ما است و چگونگی حرکات هماهنگ موضوع تحقیق‌های گسترده‌ای بوده است [۱]-[۶]. از رویکردهای کنترل حرکت بازوی انسان می‌توان به استفاده از مدل‌های توصیفی، مدل‌های کامل (شامل مدل‌های دینامیکی و مدل‌های تصادفی)، مدل‌های اجرایی حرکت و مدل‌های دینامیک مفاصل انسان اشاره کرد [۱]. هدف مدل‌های توصیفی، توصیف رفتار ظاهری حرکت انسان است. از جمله روابط تجربی این مدل‌ها که در شناسایی حرکات بازوی انسان مورد استفاده قرار می‌گیرد می‌توان به قانون فیت، پروفایل سرعت زنگوله‌وار در حرکات مستقیم و قانون ۲/۳ توان اشاده کرد. به رغم سادگی مدل‌های توصیفی، این مدل‌ها به علت در نظر نگرفتن جنبه‌های دینامیکی حرکت مورد انتقاد قرار می‌گیرند. از سوی دیگر مدل‌های دینامیکی قادر به پیش‌بینی برخی آزمایشات نبوده و در نتیجه رضایت-بخش نیستند. نشان داده شده است هنگامیکه فردی دست خود را بین موقعیت اولیه و نهایی حرکت می‌دهد، ممکن است مسیر دست خود را به منظور اصلاح خط‌سیر، با استفاده از فیدبک‌های بصری تغییر دهد [۱]. مدل‌های کینماتیک این تطبیق‌پذیری^۳ را پیش‌بینی می‌کنند در حالی که مدل‌های دینامیکی قادر به انجام چنین کاری نیستند زیرا محدوده کاری مدل‌های دینامیکی در خط‌سیر تنها به موقعیت ابتداء و انتهای حرکت بستگی دارد و موقعیت میانی در نظر گرفته نمی‌شود [۱]. هاریس و ولپرت مفهوم متفاوتی از طراحی حرکت را بیان می‌کنند که ادغامی از مفاهیم دینامیک و کینماتیک می‌باشد [۱]. سجنوسکی و همکاران، [۱]، به برخی از جنبه‌های مهم نظریه هاریس و ولپرت تحت عنوان حداقل دقت اشاره می‌کنند و بیان می‌دارند که مدل‌های مبتنی بر حداقل بهره‌وری و حرکت بدون نوسان، نمی‌توانند توصیف خوبی از خود تطبیق‌پذیری سیستم عصبی دربرداشته باشند. لی و همکارانش یک روش کلی برای کنترل پیش‌خوردی سلسه‌مراتبی ارائه دادند که می‌تواند برای مشکلات کنترل حرکت دست مورد استفاده قرار گیرد [۷]. مدل دینامیکی ارائه شده توسط لی و همکارانش، یک

¹Spinal Cord Injury(SCI)

²Functional Electrical Stimulation (FES)

³adaptation

⁴Basal ganglia

الگوی فعالیت عضلانی فرد سالم با سیگنال خروجی کنترل کننده توسعه یافته،

مقایسه شد.

۱-۱-۲- عملگر

عملگر یک کنترل کننده با عملکرد ریدیانی بالا است، که دارای ساختار شبکه عصبی-فازی می‌باشد^[۱۱]. سیگنال یادگیری تقویتی بصورت معادله (۱) بیان می‌شود، که از روش E-greedy به عنوان خط مشی عملگر به منظور تصمیم-گیری رفتار استفاده شده است.

$$u_r = \begin{cases} \sum_{j=1}^R w_j^{act} b_j^{act} \text{ probability } 1 - \epsilon \\ \sum_{j=1}^R w_j^{act} b_j^{act} + n \text{ probability } \epsilon \end{cases} \quad (1)$$

R تعداد نرون‌های لایه میانی عملگر، w_j^{act} وزن از نرون لایه میانی زام به خروجی عملگر، n نویز برای یافتن سیگنال بهینه، b_j^{act} سازگاری نرون و نرخ جستجو که در معادله (۲) آمده است.

$$\epsilon = \frac{\alpha_\epsilon |e_t|}{1 + \exp(t - 10)} \quad (2)$$

که e_t خطای خروجی سیستم و α_ϵ ضریب ثابت است. سازگاری گره از حاصلضربتابع عضویت گوسین بدست می‌آید، که بصورت معادله (۳) بیان می‌شود.

$$b_j^{act} = \prod_{i=1}^I \exp \left\{ - \frac{(e^{(i-1)} - c)^2}{(\sigma)^2 + \bar{\omega}} \right\} \quad (3)$$

که c و σ مرکز و عرض تابع پایه لایه ورودی آنم برای نرون زام لایه میانی، $e^{(i-1)}$ خطای خروجی سیستم مربوط به لایه ورودی آنم می‌باشدند و $\bar{\omega}$ ثابت مثبتی است که از صفر شدن عرض تابع پایه جلوگیری می‌کند.

ساختار عملگر خودسازمانده است، به این صورت که تعداد نرون‌های لایه میانی شبکه عصبی بصورت خودکار اضافه و کم می‌شود تا ساختار ما بصورت موثر عمل نماید. مکانیزم خودسازمانده مشابه کار چنگک [۱۹] می‌باشد. افزودن نرون به این صورت انجام می‌گیرد که:

با فرض اینکه مجموعه سازگار از نرون b_j^{act} داریم، بزرگترین خروجی نرون‌ها برابر است با:

$$\Gamma_{\max} = \max(b_j^{act}) \quad j = 1, 2, \dots, R(t) \quad (4)$$

که $R(t)$ تعداد نرون‌ها در زمان t است. شرایط اضافه شدن نرون جدید در لایه میانی به اینصورت است که:

$$\Gamma_{\max}(t) \leq \Gamma_{\text{th}} \quad (5)$$

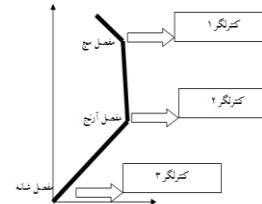
آستانه افزودن به منظور اضافه کردن نرون است. اگر معادله (۵) برآورده شود، نرون جدید به لایه میانی اضافه می‌شود. هر نرون دارای یک شاخص اهمیت I_i است [۱۹] که نرون با استفاده از آن حذف می‌شود. به-

روزرسانی به اینصورت انجام می‌شود که:

$$I_j(t+1) = \begin{cases} I_j(t) & \text{if } b_j^{act} \geq P_{\text{th}} \\ I_j(t). \exp(-\beta) & \text{if } b_j^{act} < P_{\text{th}} \end{cases} \quad (6)$$

۲- ارائه راهکار کنترل

حرکت دست به سمت هدف، بوسیله ایجاد گشتاور حول سه مفصل مچ، آرنج و شانه ایجاد می‌شود. لذا برای کنترل چنین حرکتی دو راه حل وجود دارد. راه حل اول این است که، به از یک کنترل کننده چند ورودی-چند خروجی استفاده شود. اما طراحی این کنترل کننده مشکل بوده و به لحاظ پیاده‌سازی بار محاسباتی بالایی دارد، از طرفی تضمین پایداری چنین ساختاری بسیار مشکل است. راه حل دوم این است که یک استراتژی کنترل غیرمتمن کر توسعه داده شود، بطوریکه برای کنترل حرکت هر مفصل از یک کنترل کننده مجزا استفاده شود، که ساختار کلی آن در شکل (۱) نشان داده شده است. در ساختار کنترلی غیر-متمنکر، کوپلینگ‌های بین زیر سیستم‌ها، به عنوان دینامیکهای مدل نشده یا اغتشاش برای هر زیر سیستم در نظر گرفته می‌شود.

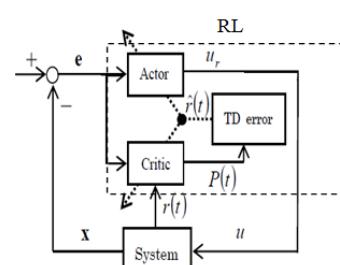


شکل ۱: سیستم کنترل غیرمتمنکر به منظور کنترل حرکت بازوی دست

ابتدا کنترل کننده مبتنی بر یادگیری تقویتی (RL) شامل دو قسم عملگر^۵ و ارزیاب^۶ در جایگاه کنترل کننده قرار گرفت. پس از بررسی نتایج بدست آمده به منظور بهبود کارایی، کنترل کننده PID به ساختار کنترلی اضافه شد. سپس کنترل کننده HTC در جهت تضمین پایداری سیستم و رسیدن به نتیجه مطلوبتر به ساختار کنترلی اضافه گردید.

۲-۱- کنترل کننده عملگر- ارزیاب

ساختار کلاسیک کنترل کننده یادگیری تقویتی شامل دو بخش است: ۱) عملگر^۷؛ ۲) ارزیاب. عملگر سیگنال مطلوب را تولید کرده و ارزیاب وظیفه ارزیابی عملکرد شبکه عملگر را بر عهده دارد، به عبارتی عملگر بر اساس خروجی‌های ارزیاب تطبیق می‌یابد^[۱۸]. شکل (۲) ساختار کلی کنترل کننده مبتنی بر یادگیری تقویتی را نشان می‌دهد.



شکل ۲: ساختار کنترل کننده مبتنی بر یادگیری تقویتی

⁵Actor

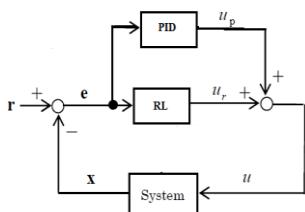
⁶Critic

۲-۲- کنترل کننده مبتنی بر یادگیری تقویتی و PID

در ساختار نشان داده شده در شکل (۱)، از ترکیب کنترل کننده مبتنی بر یادگیری تقویتی و کنترل کننده پسخوردنی به منظور کنترل حرکت دست و بهبود کارایی سیستم استفاده شده است. کنترل کننده تابعی-انتگرالی-مشتقی کنترل کننده‌ای مبتنی بر فیدبک از خروجی می‌باشد. این کنترل کننده یکی از استراتژی‌های متداول در کنترل سیستم‌ها است. در ساختار پیشنهادی، از یک کنترل کننده PID، در کنار عملگر ارزیاب استفاده شده است. با توجه به اینکه مدل بازو دارای سه مفصل بوده است، برای کنترل هر مفصل از یک کنترل کننده پاداش p را محاسبه کرده تا خطای اختلاف زمانی^۷ را به حداقل برساند. پاداش کننده PID عبارتست از:

$$G_c(s) = K_p + \frac{K_I}{s} + K_D s \quad (۱۴)$$

که در آن K_p گین تابعی، K_I گین انتگرالی و K_D گین مشتقی است. تحقق فیزیکی این کنترل کننده توسط یک مدار الکتریکی امکان‌پذیر است. به منظور کنترل سیستم، از PID به عنوان جبرانساز در کنار راهکار کنترلی مبتنی بر روش یادگیری تقویتی (RLPID) استفاده می‌شود.



شکل ۳: کنترل کننده مبتنی بر یادگیری تقویتی و کنترل کننده PID

۳-۲- کنترل کننده مبتنی بر یادگیری تقویتی و HTC و PID

به منظور بهبود پایداری سیستم کنترل علاوه بر کنترل کننده PID یک کنترل کننده مقاوم به ترکیب افزوده شده است. این کنترل کننده مقاوم که از نوع H_{∞} است، HTC^۹ (جبرانساز با دقت دنبال کنندگی بالا) نامیده می‌شود. ساختار این کنترل کننده در شکل (۴) آمده است. رابطه (۱۵) سیگنال کنترلی خروجی HTC را نشان می‌دهد. هدف عملگر توسط کنترل کننده مقاوم مشاهده و خروجی u_h مربوط به HTC بگونه‌ای طراحی می‌شود که عملکرد ردیابی H_{∞} را تضمین نماید.

$$u_h = \frac{1}{8\tau^2} e^T P B, \quad B = [0 \ 0 \dots 0 \ 1]^T \quad (۱۵)$$

τ ثابت است.

$$\delta = 2\tau \frac{g}{\sqrt{g}}$$

δ ثابت میرایی است. از آنجا که g ناشناخته است، فرض می‌کنیم:

$$g_{low} < g < g_{up}$$

در نتیجه داریم:

$$\delta = 2\tau \frac{g_{up}}{\sqrt{g_{low}}} \geq 2\tau \frac{g}{\sqrt{g}}$$

تا زمانیکه یادگیری به شکل مطلوبی اتفاق نیافتد، HTC نقش موثرتری در کنترل سیستم بر عهده دارد، اما پس از اینکه یادگیری به سطح

β شاخص طراحی و P_{th} آستانه تغییر است. اگر شاخص اهمیت نرون از آستانه تغییر کمتر باشد، در نتیجه این مقدار کاهش می‌یابد. مقدار اولیه شاخص اهمیت، را برابر یک انتخاب شده است.

$$I_j \leq I_{th} \quad (۷)$$

که I_{th} آستانه حذف است. اگر معادله (۷) برای نرون برآورده شود، یک نرون نامناسب تشخیص داده شده و حذف می‌شود.

۲-۱-۲- ارزیاب

شبکه عصبی مصنوعی قسمت ارزیاب را شکل داده است[۱۱]. ارزیاب پاداش p را محاسبه کرده تا خطای اختلاف زمانی^۷ را به حداقل برساند. پاداش مورد نظر برابر است با:

$$P(t) = \sum_{j=1}^J w_j^{cri} b_j^{cri} \quad (۸)$$

که در آن J تعداد نرون لایه میانی ارزیاب و w_j^{cri} وزن نرون زام لایه میانی به خروجی است و b_j^{cri} با استفاده از تابع هذلولی مماسی (tanh) تحقق می‌یابد. ایندهی مرکزی و بدیع یادگیری تقویتی بدون شک یادگیری تفاضل زمانی خواهد بود[۲۰]. در اینجا یادگیری عملگر ارزیاب با استفاده از خطای تفاضل زمانی انجام می‌گیرد، که بصورت معادله (۹) بیان می‌شود:

$$\hat{r}(t) = r(t) + \gamma \max P(t+1) - P(t) \quad (۹)$$

که r پاداش و γ نرخ نزول ($1 < \gamma < 0$) می‌باشد.

$$r = \alpha_\theta e_t + \alpha_\theta \dot{e}_t \quad (۱۰)$$

α_θ و α_θ مقادیر ثابت و e_t خطای خروجی سیستم هستند. ویژگی خروجی برای دریافت بیشترین پاداش، گین عملگر است. به عبارتی یادگیری عملگر به منظور افزایش پاداش و یادگیری ارزیاب به منظور کاهش خطای TD انجام می‌گیرد. در نتیجه در این مقاله الگوریتم خطای پس انتشار خطای به منظور یادگیری شبکه عصبی استفاده می‌شود. در واقع یادگیری عملگر براساس معادله (۱۱) و یادگیری ارزیاب برطبق معادله (۱۲) خواهد بود.

$$\Delta w_j^{act} = \eta_w^{act} \frac{\partial u}{\partial w_j^{act}} \hat{r} \quad (۱۱)$$

$$\Delta w_j^{cri} = -\eta_w^{cri} \frac{\partial P}{\partial w_j^{cri}} \frac{\partial}{\partial P} \left(\frac{1}{2} \hat{r}^2 \right) \quad (۱۲)$$

η_w^{act} و η_w^{cri} بترتیب نرخ یادگیری ارزیاب و عملگر و \hat{r} سیگنال خطای می-باشند.

از روش پروجکشن^۸ برای محدود کردن w_j^{act} استفاده می‌کنیم، به این معنا که یک قانون به روزرسانی برای w_j^{act} بصورت زیر تعریف می‌کنیم:

$$w_j^{act} = \begin{cases} w_j^{act} + \Delta w_j^{act} if |w_j^{act} + \Delta w_j^{act}| < D \\ w_j^{act} otherwise \end{cases} \quad (۱۳)$$

که محدوده موثر وزن می‌باشد. در واقع اگر قدر مطلق وزن پس از یادگیری از D کمتر باشد، به روز رسانی w_j^{act} انجام می‌گیرد.

^۹ H_{∞} Tracking Compensator

⁷ Temporal Difference error (TD)

⁸ projection

جایگزینی معادله (۲۲) در معادله (۲۱) داریم:

$$\dot{e} = \alpha e + B_l(u^* - u_r - u_h - u_p) \quad (23)$$

خطای تقریب سیگنال یادگیری تقویتی، کنترل کننده PID و ورودی بهینه

به صورت زیر می‌باشد:

$$\varepsilon_u = u^* - u_r - u_p \quad (24)$$

با جایگذاری معادله (۲۴) در (۲۳) داریم:

$$\dot{e} = \alpha e + B_l(\varepsilon_u - u_h) \quad (25)$$

برطبق معادله لیاپانوف، تطبیق به این صورت است:

$$V = e^T P e \quad (26)$$

با مشتق گیری از دو طرف معادله (۲۶) نسبت به زمان، داریم:

$$\dot{V} = \dot{e}^T P e + e^T P \dot{e} \quad (27)$$

با استفاده از معادلات (۱۷) و (۲۳):

$$\dot{V} = -e^T Q e - 2e^T P B_l u_h + 2e^T P B_l \varepsilon_u \quad (28)$$

با قراردادن معادله (۱۵) در (۲۸) داریم:

$$\begin{aligned} \dot{V} = -e^T Q e - & \left\{ \left(\frac{\sqrt{g}}{2\tau} \right) e^T P B - 2\tau \left(\frac{g}{\sqrt{g}} \right) \varepsilon_u \right\} \\ & + \left(2\tau \frac{g}{\sqrt{g}} \right) \varepsilon_u^T \left(2\tau \frac{g}{\sqrt{g}} \right) \varepsilon_u^2 \end{aligned} \quad (29)$$

در نتیجه:

$$\dot{V} \leq -e^T Q e + \left(2\tau \frac{g}{\sqrt{g}} \right) \varepsilon_u^T \left(2\tau \frac{g}{\sqrt{g}} \right) \varepsilon_u \quad (30)$$

از دو طرف معادله (۳۰) انتگرال می‌گیریم:

$$v(t_f) - v(0) \leq - \int_0^{t_f} e^T Q e dt + \left(2\tau \frac{g}{\sqrt{g}} \right)^2 \int_0^{t_f} \varepsilon_u^T \varepsilon_u dt \quad (31)$$

از آنجا که $v(0) = e^T(0)Pe(0)$ و $v(t_f) = e^T(t_f)Pe(t_f)$ داریم:

$$\int_0^{t_f} e^T Q e dt \leq e^T(0)Pe(0) + \left(2\tau \frac{g}{\sqrt{g}} \right)^2 \int_0^{t_f} \varepsilon_u^T \varepsilon_u dt \quad (32)$$

که زمان کنترل نهایی می‌باشد. همانطور که از معادله (۳۲) قابل مشاهده است، انتگرال خطای ردیابی e توسط مقادیر بدست آمده توسط ضرب خطا تقریب ε_u و ثابت میرایی δ محدود شده است. به عبارتی اگر خطای تقریب و ثابت میرایی هردو محدود باشد، خطای ردیابی نیز محدود است، در نتیجه پایداری سیستم کنترل اثبات می‌شود.

در نتیجه HTC بصورت معادله (۳۳) بیان می‌شود:

$$\int_0^{t_f} e^T Q e \leq e^T(0)Pe(0) + \delta^2 \int_0^{t_f} \varepsilon_u^T \varepsilon_u dt \quad (33)$$

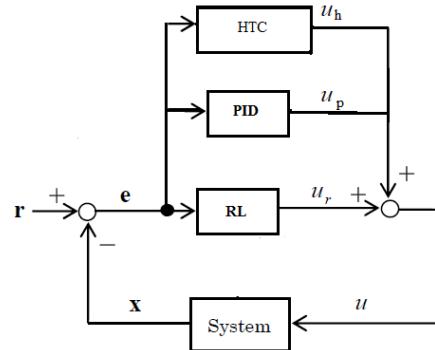
که t_f زمان کنترل نهایی، ε_u مجموع عدم قطعیت خطای تقریب و اختشاشات در زمان t و δ ثابت مثبت غیر صفر می‌باشد [۱۱].

۳- مدل بازوی مجازی

مطالعات شیوه‌سازی مربوط به این تحقیق روی مدلی از بازوی دست انسان شامل سه مفصل انجام شده‌اند. شکل (۵) ساختار کلی این مدل را نشان می‌دهد [۶]. معادله دینامیکی بازوی سه مفصلی انسان که در صفحه افقی حرکت می‌کند، توسط معادله (۳۴) بیان می‌شود.

$$\begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix} \begin{bmatrix} \ddot{\theta}_1 \\ \ddot{\theta}_2 \\ \ddot{\theta}_3 \end{bmatrix} + \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} + \begin{bmatrix} D_1 & 0 & 0 \\ 0 & D_2 & 0 \\ 0 & 0 & D_3 \end{bmatrix} \begin{bmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \\ \dot{\theta}_3 \end{bmatrix} = M\ddot{\theta} + h + D\dot{\theta} \quad (34)$$

مطلوبی رسید، هر دو کنترل کننده مبتنی بر یادگیری تقویتی و HTC در کنترل عملکرد حرکتی موثر خواهد بود. خروجی H_∞ با استفاده از حل P در معادله لیاپانوف ایجاد می‌شود.



شکل ۴: کنترل کننده مبتنی بر یادگیری تقویتی و کنترل کننده PID و H_∞ در روش RLPID اثبات تئوری برای تضمین پایداری سیستم کنترل وجود ندارد. اما در حضور کنترل کننده HTC می‌توان با تعریفتابع لیاپانوف پایداری سیستم کنترل حلقه بسته را اثبات کرد. در زیر بخش بعد پایداری سیستم اثبات می‌شود.

۱-۳-۲- آنالیز پایداری سیستم کنترل

سیستم غیرخطی مرتبه اول را بصورت زیر در نظر می‌گیریم:

$$\dot{x} = A(x) + B(x)u \quad (16)$$

که x بردار حالت سیستم، f تابع پیوسته محدود ناشناخته، g تابع پیوسته مثبت محدود ناشناخته و u ورودی کنترل به سیستم می‌باشد.

تابع لیاپانوف را با در نظر گرفتن ماتریس قطری A که قطر اصلی آن شامل بهره پس خور k می‌باشد ($A = k^T e$)، a که خطای دنبال کنندگی، ماتریس مثبت معین P و ماتریس مثبت معین متعامد Q ، که بصورت قراردادی تعیین شده، حل می‌نماییم با این شرط که پایداری هرویتز را برآورده نماید [۱۱].

$$A^T P + P A = -Q \quad (17)$$

اگر A و B شناخته شده باشند، ورودی بهینه سیستم برابر است با:

$$u^* = \frac{1}{B} (-A + \dot{r} + k^T e) \quad (18)$$

که \dot{r} مشتق اول سیگنال مرجع و e خطای ردیابی ($e = r - x$) است که باید صفر شود.

اگر سیستم (۱۶) را بصورت زیر در نظر می‌گیریم:

$$\dot{x} = A + Bu + Bu^* - Bu^* \quad (19)$$

مشتق خطای ردیابی برابر است با:

$$\begin{aligned} \dot{e} = \dot{r} - \dot{x} = & (A + gu^* - k^T e) \\ & - (A + Bu + Bu^* - Bu^*) \\ & = -k^T e + B(u^* - u) \end{aligned} \quad (20)$$

معادله فضای حالت مشتق خطای ردیابی بصورت زیر می‌باشد:

$$\dot{e} = \alpha e + B_l(u^* - u) \quad (21)$$

ورودی کنترلی سیستم مبتنی بر RLPIDHTC بصورت معادله (۲۲) بیان می‌شود:

$$u = u_r + u_h + u_p \quad (22)$$

و u_r ، u_h ، u_p ، بترتیب، سیگنال خروجی HTC و PID و RL هستند. با

I بردار 3×3 واحد است. جدول (۱) خلاصه‌ای از پارامترهای فیزیکی بازو را نشان می‌دهد.

که M ماتریس اینرسی، h بردار نیروی گریز از مرکز و جانب مرکز و D بردار نیروی ویسکوزیته می‌باشد.

جدول ۱: پارامترهای دینامیکی بازو سه مفصل [۶]

مُج	آرنج	شانه	پارامترها
۰.۱۳۰	۰.۲۷۵	۰.۲۷۶	Li(m)
۰.۰۴۳	۰.۰۸۶	۰.۱۴۱	Lci(m)
۰.۴۸۱	۰.۹۵۱	۱.۷۴۸	mi(kg)
۰.۰۰۵	۰.۰۰۴۱	۰.۰۱۲۷	Ii(kg.m ²)
۰.۱۹۰	۰.۲۹۰	۰.۳۷۰	Di(Nm.s/rad)

۴- مطالعات شبیه‌سازی

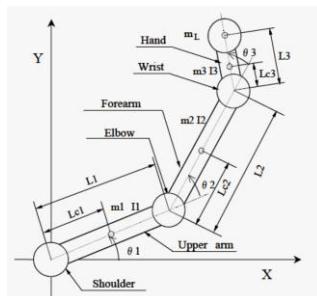
مطالعات شبیه‌سازی مربوطه در محیط سیمولینک مطلب با زمان نمونه-برداری ۰/۰۴ ثانیه انجام شدند. در مطالعات شبیه‌سازی، قابلیت راهکارهای کنترلی مورد مطالعه، در رساندن موقعیت مفاصل بازوی دست از موقعیت‌های مختلف به موقعیت تعادلی [0,0,0]، مورد ارزیابی قرار گرفته است. همچنین کنترل کننده‌ها در حضور انگشت‌شان، تغییر پارامترهای سیستم و نویز اندازه‌گیری مورد ارزیابی قرار گرفته‌اند. در ادامه نتایج مطالعات شبیه‌سازی به تفصیل ارائه خواهد شد.

۴-۱- کنترل کننده RL

همانطور که پیشتر عنوان شد، به منظور کنترل حرکت بازو از سه کنترل-کننده مجزا برای هر مفصل استفاده می‌شود. در این قسمت سه کنترل کننده مبتنی بر یادگیری تقویتی (RL) به منظور کنترل حرکت مفاصل موردن استفاده قرار گرفته است. پارامترهای استفاده شده به منظور شبیه‌سازی کنترل کننده RL در جدول (۲) آمده است. کنترل کننده RL در چند اپیزود مختلف بر سیستم اعمال شد. اما نتایج بدست آمده در تمام اپیزودها مشابه بودند. لذا در شکل‌های (۶) تا (۸) تغییرات زوایای مفاصل و سیگنال خروجی کنترلی هر سه مفصل به ترتیب برای شرایط اولیه [۱,۱,۱]، [۵,۵,۵] و [۵-۵,۵-۵] در اپیزود اول نشان داده شده‌اند.

جدول ۲: پارامترهای استفاده شده در شبیه‌سازی سیستم کنترل کننده RL

k	{1,5}	γ	0.95
Q	diag{10,10}	α_θ	5
δ	0.3	$\alpha_{\dot{\theta}}$	1
g _{low}	0.6	α_ε	0.1
g _{up}	1.5	η_w^{cri}	0.01
I	2	η_w^{act}	1
J	5	D	500
Γ_{th}	0.3	n	Random(-10-10)
σ_c	0.2	$\bar{\omega}$	0.001
I _{th}	0.01		



شکل ۵: مدل بازوی ۳-مفصل انسان [۶]

لينک‌های ۱، ۲ و ۳ به ترتیب متاظر با بالای بازو، ساعد و مج دست هستند و مفاصل ۱، ۲ و ۳ به ترتیب مربوط به مفاصل شانه، آرنج و ساعد می‌باشند. m_i و m_L به ترتیب جرم لينک i و جرم (= جسم گرفته شده i) I_i و I_L به ترتیب طول، فاصله مرکز جرم i تا مفصل اتم ($i=1,2,3$) است. مزیت این روش این است که مسیر باز تولید شده با حالت تجربی، بدون توجه به موقعیت هدف، همخوانی دارد [۶].

۳-۱- مکانیزم فریزسازی

در مدل بازوی مجازی به منظور لحاظ کردن اثر فعالیت عضلات آنتاگونیستی از مکانیزم فریزسازی استفاده شده است. سفتی مفاصل دست در اثر پدیده فریزسازی، که در آن عضلات آگونیست و آنتاگونیست همزمان متقاضی می‌شوند، افزایش می‌یابد [۶]. در این مکانیزم یک مولفه گشتاور مطابق رابطه (۳۵) به گشتاور حاصل از افزوده شده است، چراکه در اثر گرفتن جسم، بیشترین گشتاور حاصل از فعالیت عضلات آنتاگونیستی به مج اعمال می‌شود. گشتاور پسخوردی حاصل از مکانیزم فریزسازی $^3\tau$ در مفصل مج دست بصورت زیر بیان می‌شود [۶]:

$$\tau_3^* = -k_a \ddot{\theta}_3 - k_v \dot{\theta}_3 - k_p \theta_3 \quad (35)$$

k_p و k_v گین‌های فیدبک هستند. با لحاظ مکانیزم فریزسازی معادله

دینامیکی بازو به صورت معادله (۳۶) تغییر می‌یابد.

$$\begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 + \tau_3^* \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix} \begin{bmatrix} \ddot{\theta}_1 \\ \ddot{\theta}_2 \\ \ddot{\theta}_3 \end{bmatrix} + \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} + \begin{bmatrix} D_1 & 0 & 0 \\ 0 & D_2 & 0 \\ 0 & 0 & D_3 \end{bmatrix} \begin{bmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \\ \dot{\theta}_3 \end{bmatrix} \quad (36)$$

ماتریسی اینرسی، h بردار نیروی جانب مرکز و گریز از مرکز و D بردار نیروی ویسکوزیته می‌باشد. با این فرض که حرکت بازو در صفحه افقی و در

ارتفاع شانه انجام می‌شود، معادلات حالت زیر بدست می‌آید:

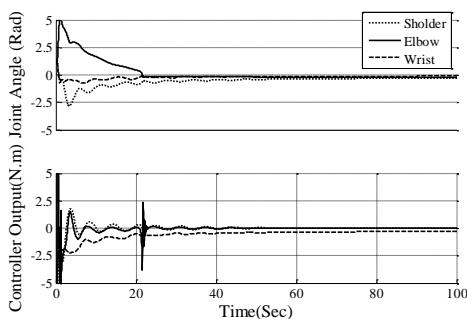
$$\dot{x} = \begin{bmatrix} \dot{\theta} \\ -\bar{M}^{-1}(\bar{h} + \bar{D}\dot{\theta}) + \bar{M}^{-1}\tau \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u = A(x) + Bu \quad (37)$$

$$\theta = [\theta_1, \theta_2, \theta_3]^T, \tau = [\tau_1, \tau_2, \tau_3]^T, u = [\dot{t}_1, \dot{t}_2, \dot{t}_3]^T \\ x = [\theta_1, \theta_2, \theta_3, \dot{\theta}_1, \dot{\theta}_2, \dot{\theta}_3, t_1, t_2, t_3]^T$$

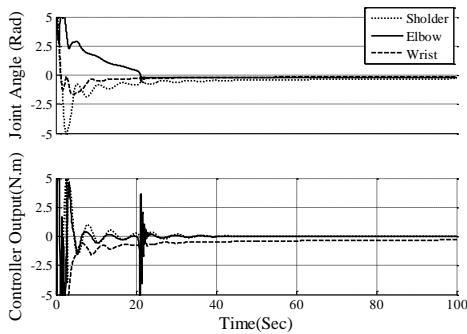
¹⁰handle

¹¹Center of mass

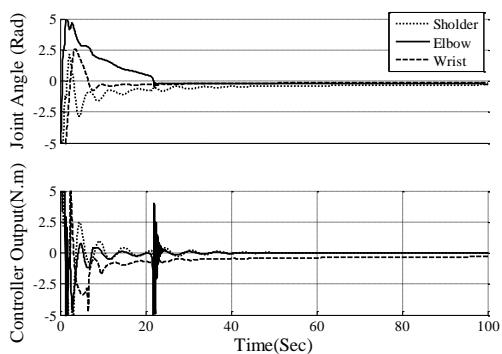
k_{p1} و k_{d1} گین‌های مربوط به کنترل کننده PID مفصل شانه، k_{p2} و k_{d2} گین‌های مربوط به کنترل کننده PID مفصل آرنج و k_{p3} و k_{d3} گین‌های مربوط به کنترل کننده PID مفصل مچ دست می‌باشند، که ساختار کلی آن در شکل (۱) نشان داده شده است. میانگین مربعات خطأ و همچنین خطای حالت ماندگار زوایای مفاصل دست با شرایط اولیه متفاوت، به ترتیب، در جدول‌های (۳) و (۴) آمده است. در جدول (۵) حداکثر فروجهش برای شرایط اولیه [۱,۱,۱] و حداقل فروجهش برای شرایط اولیه [۵,۵,۵] آمده است. زمان صعود برای شرایط اولیه [-۵,-۵,-۵] و زمان نزول برای شرایط اولیه [۱,۱,۱] و [۵,۵,۵] در جدول (۶) محاسبه شده است.



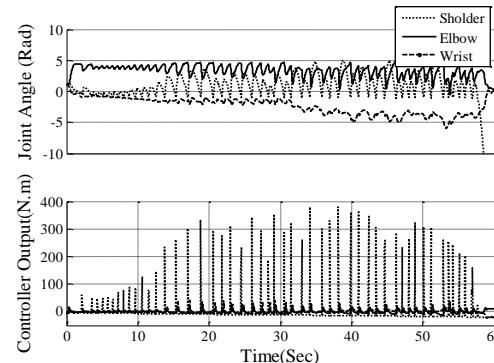
شکل ۹: زوایای مفاصل و خروجی کنترل کننده RLPID برای سه مفصل شانه، آرنج و دست با شرایط اولیه $\theta_0 = [1,1,1]$



شکل ۱۰: زوایای مفاصل و خروجی کنترل کننده RLPID برای سه مفصل شانه، آرنج و دست با شرایط اولیه $\theta_0 = [5,5,5]$

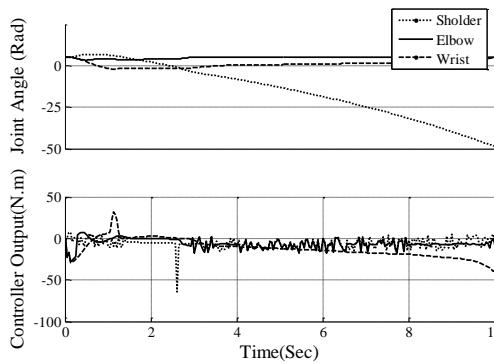


شکل ۱۱: زوایای مفاصل و خروجی کنترل کننده RLPID برای سه مفصل شانه، آرنج و دست با شرایط اولیه $\theta_0 = [-5,-5,-5]$



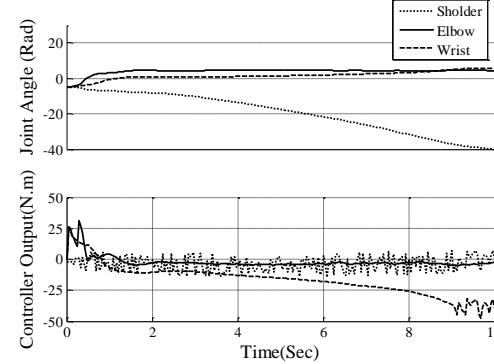
شکل ۶: زوایای مفاصل و خروجی کنترل کننده RL برای سه مفصل شانه

$$\text{آرنج و دست با شرایط اولیه } \theta_0 = [1,1,1]$$



شکل ۷: زوایای مفاصل و خروجی کنترل کننده RL برای سه مفصل شانه

$$\text{آرنج و دست با شرایط اولیه } \theta_0 = [5,5,5]$$



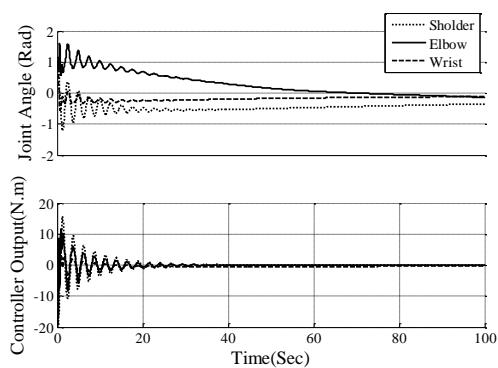
شکل ۸: زوایای مفاصل و خروجی کنترل کننده RL برای سه مفصل شانه

$$\text{آرنج و دست با شرایط اولیه } \theta_0 = [-5,-5,-5]$$

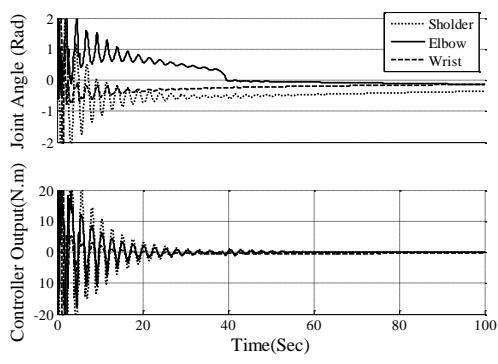
۴-۲- کنترل کننده RLPID

همانطور که از نتایج کنترل کننده RL قابل مشاهده است، این کنترل کننده ناپایدار است. به منظور بهبود کارایی و پایداری سیستم از کنترل کننده PID در کار کنترل کننده RL استفاده خواهد شد. به منظور بررسی عملکرد کنترل کننده یادگیری تقویتی در کار کنترل کننده PID، از ساختار شکل (۳) استفاده می‌شود. نتایج حاصل از اعمال همزمان کنترل کننده‌های یادگیری تقویتی و PID به مفاصل دست و همچنین خروجی‌های کنترلی با شرایط اولیه [۱,۱,۱]، [۵,۵,۵] و [-۵,-۵,-۵] به ترتیب در شکل‌های (۹) تا (۱۱) آمده است. پارامترهای کنترل-کننده‌های PID بصورت زیر تنظیم شده‌اند:

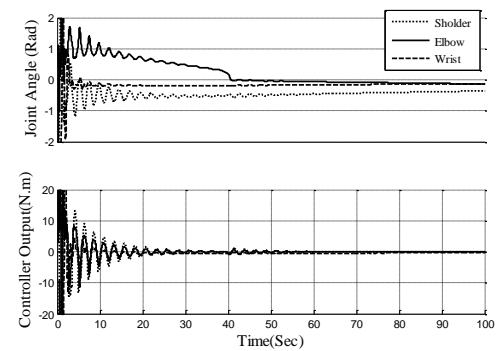
$$\begin{cases} K_{p1} = 2, K_{i1} = 0.5, K_{d1} = 1 \\ K_{p2} = 5, K_{i2} = 0.5, K_{d2} = 1 \\ K_{p3} = 2, K_{i3} = 1, K_{d3} = 1 \end{cases} \quad (۳۸)$$



شکل ۱۲: زوایای مفاصل و خروجی کنترل کننده RLPIDHTC برای سه مفصل شانه، آرنج و دست با شرایط اولیه $\theta_0 = [1,1,1]$



شکل ۱۳: زوایای مفاصل و خروجی کنترل کننده RLPIDHTC برای سه مفصل شانه، آرنج و دست با شرایط اولیه $\theta_0 = [5,5,5]$



شکل ۱۴: زوایای مفاصل و خروجی کنترل کننده RLPIDHTC برای سه مفصل شانه، آرنج و دست با شرایط اولیه $\theta_0 = [-5, -5, -5]$

جدول ۳: میانگین مربعات خطای زوایای مفاصل دست مربوط به کنترل کننده

RLPID		
[-5,-5,-5]	[5,5,5]	[1,1,1]
۰.۶۷۴۸°	۱.۰۴۴۶°	۰.۴۸۹۶°
۱.۱۸۵۶°	۱.۱۲۲۱°	۱.۰۵۲۷°
۰.۶۵۹۳°	۰.۳۰۰۲°	۰.۰۶۲۵°
میانگین		

جدول ۴: خطای حالت ماندگار زوایای مفاصل دست مربوط به کنترل-

RLPID		
کننده		
[-5,-5,-5]	[5,5,5]	[1,1,1]
۰.۶۰۵۲°	۰.۸۵۸۰°	۰.۴۲۵۵°
۱.۰۵۶۴°	۱.۰۲۵۹°	۰.۹۹۱۵°
۰.۷۸۸۳°	۰.۵۱۲۰°	۰.۱۵۶۱°
میانگین		

جدول ۵: حداکثر فرجهش (فروجهش) زوایای مفاصل دست مربوط به

RLPID		
کننده		
[-5,-5,-5]	[5,5,5]	[1,1,1]
۲.۱۱۱۱°	-۵.۱۱۹۸°	-۲.۷۸۶۴°
۶.۰۵۹۶°	-۰.۳۶۹۲°	-۰.۲۴۶۷°
۲.۵۸۶۴°	-۱.۶۷۹۱°	-۰.۷۴۸۴°
میانگین		

جدول ۶: زمان صعود (نزول) زوایای مفاصل دست مربوط به کنترل کننده

RLPID		
[-5,-5,-5]	[5,5,5]	[1,1,1]
۱۵.۴	۱۴.۶۴	۵۲.۵۶
۱۴.۵۲	۱۳.۶	۲۱.۸۴
۵.۶	۷.۱۲	۳۶.۵۲
میانگین		

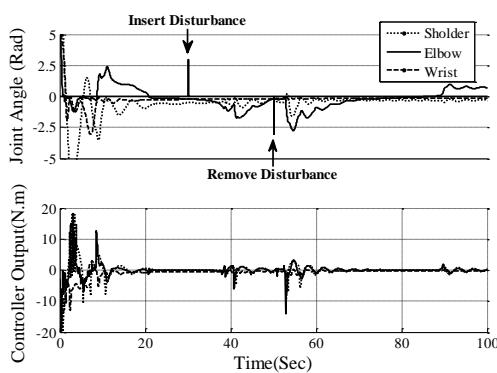
۴-۳- کنترل کننده RLPIDHTC

همانطور که پیشتر مشاهده شد کارایی کنترل کننده RL مطلوب نبود. در ادامه از کنترل کننده RLPID به منظور کنترل سیستم استفاده شد. با وجود کارایی قابل قبول، به منظور بهبود شرایط پایداری سیستم کنترل، یک کنترل کننده مقاوم HTC به سیستم کنترل مبتنی بر RLPID افزوده شد که با عنوان کنترل کننده RLPIDHTC معروف شد. نتایج حاصل از اعمال همزمان کنترل کننده‌های یادگیری تقویتی، HTC و PID به منظور رساندن موقعیت مفاصل دست به موقعیت تعادلی $[0,0,0]$ ، از شرایط اولیه $[1,1,1]$ ، $[5,5,5]$ و $[-5,-5,-5]$ برای هر سه مفصل در شکل‌های (۱۲) تا (۱۴) آمده است. در جدول‌های (۷) و (۸) نتایج مربوط به میانگین مربعات خطأ و خطای حالت ماندگار حاصل از اعمال ساختار کنترلی RLPIDHTC آورده شده‌اند. در جدول (۹) حداکثر فرجهش برای شرایط اولیه $[-5,-5,-5]$ و حداقل فرجهش برای شرایط اولیه $[1,1,1]$ و $[5,5,5]$ آمده است. زمان صعود برای شرایط اولیه $[-5,-5,-5]$ و زمان نزول برای شرایط اولیه $[1,1,1]$ و $[5,5,5]$ در جدول (۱۰) محاسبه شده است.

جدول ۷: میانگین مربعات خطای زوایای مفاصل دست مربوط به کنترل کننده

RLPIDHTC

[-5,-5,-5]	[5,5,5]	[1,1,1]	شانه
۰.۳۷۶۲°	۰.۵۱۹۵°	۰.۲۴۰۹°	
۰.۳۱۶۱°	۰.۲۸۳۰°	۰.۲۲۶۷°	آرنج
۰.۲۵۵۴°	۰.۲۶۳۴°	۰.۰۴۰۲°	میانگین



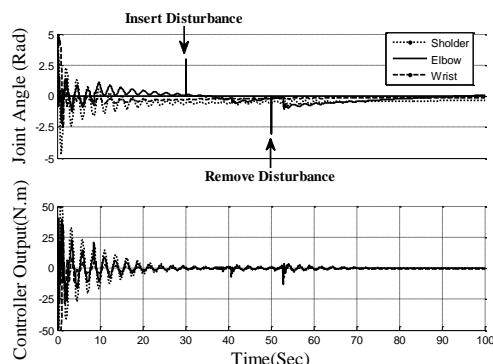
شکل ۱۰: زوایای مفاصل و خروجی کنترل کننده RLPID برای سه مفصل شانه، آرنج و دست با شرایط اولیه $\theta_0 = [5,5,5]$ در حضور اختشاش $(td=30s, 50s)$

جدول ۱۱: میانگین مربعات خطای زوایای مفاصل دست مربوط به کنترل کننده RLPID در حضور اختشاش خارجی

$[-5,-5,-5]$	$[5,5,5]$	$[1,1,1]$	
0.7675°	1.5803°	0.8246°	شانه
0.8052°	0.7594°	0.5798°	آرنج
0.3080°	0.4344°	0.0925°	مج

جدول ۱۲: خطای حالت ماندگار زوایای مفاصل دست مربوط به کنترل کننده RLPID در حضور اختشاش خارجی

$[-5,-5,-5]$	$[5,5,5]$	$[1,1,1]$	
0.6725°	1.1217°	0.7172°	شانه
0.8832°	0.8564°	0.8111°	آرنج
0.5122°	0.6213°	0.235°	مج



شکل ۱۳: زوایای مفاصل و خروجی کنترل کننده RLPIDHTC برای سه مفصل شانه، آرنج و دست با شرایط اولیه $\theta_0 = [5,5,5]$ در حضور اختشاش $(td=30s, 50s)$.

جدول ۸: خطای حالت ماندگار زوایای مفاصل دست مربوط به کنترل -

RLPIDHTC		
$[-5,-5,-5]$	$[5,5,5]$	$[1,1,1]$
0.4116°	0.5594°	0.1849°
0.4812°	0.4454°	0.3756°
0.4728°	0.4813°	0.0910°

جدول ۹: حداکثر فراجهش و فروجهش زوایای مفاصل دست مربوط به

RLPIDHTC		
$[-5,-5,-5]$	$[5,5,5]$	$[1,1,1]$
4.1682°	-4.6789°	-1.2257°
3.4015°	-0.6004°	-0.1375°
2.9116°	-2.5413°	-0.5595°

جدول ۱۰: زمان صعود (نزول) زوایای مفاصل دست مربوط به کنترل کننده

RLPIDHTC		
$[-5,-5,-5]$	$[5,5,5]$	$[1,1,1]$
4.4	10.72	81.24
12.28	14.28	46.44
2.16	1.84	38.08

۴-۴- ارزیابی در حضور اختشاش خارجی

پس از ارزیابی های اولیه ملاحظه شد که کنترل کننده های غیرمتمن کر مبتنی بر PID و RLPIDHTC در کنترل موقعیت دست، کارایی قابل قبولی داشته اند. در ادامه به منظور مقایسه دقیق تر کارایی دو کنترل کننده، قابلیت آنها در مقابله با اعمال اختشاش مکانیکی خارجی مورد بررسی قرار گرفت. به این منظور، در ثانیه 30° پالسی با دامنه 3 Nm و در ثانیه 50° پالسی با دامنه 3 Nm در مدت ۱ ثانیه به مفاصل دست اعمال شدند. عملکرد کنترل کننده RLPIDHTC بازی در نظر گرفتن موقعیت اولیه $[5,5,5]$ برای مفاصل، در شکل (۱۵) نشان داده شده است.

میانگین مربعات خطای خطا و خطای حالت ماندگار مربوط به زوایای مفاصل با شرایط اولیه مختلف در جدول های (۱۱) و (۱۲) آمده است. عملکرد کنترل کننده RLPIDHTC نیز بازی در نظر گرفتن موقعیت اولیه $[5,5,5]$ برای مفاصل، در شکل (۱۶) نشان داده شده است. میانگین مربعات خطای خطا و خطای حالت ماندگار مربوط به زوایای مفاصل با شرایط اولیه مختلف در جدول های (۱۳) و (۱۴) آمده است.

جدول ۱۵: میانگین مربعات خطای زوایای مفاصل دست مربوط به کنترل-

کننده RLPID در اثر تغییر پارامتر

[-5,-5,-5]	[5,5,5]	[1,1,1]	
۰.۶۶۵۳°	۰.۹۱۴۰°	۰.۴۷۴۱°	شانه
۱.۲۱۹۶°	۱.۰۹۱۷°	۱.۰۵۰۲°	آرنج
۰.۵۸۸۱°	۰.۳۰۶۸°	۰.۰۶۲۳°	مج

جدول ۱۳: میانگین مربعات خطای زوایای مفاصل دست مربوط به کنترل-

کننده RLPIDHTC در حضور اختشاش خارجی

[-5,-5,-5]	[5,5,5]	[1,1,1]	
۰.۴۸۳۸°	۰.۵۳۰۴°	۰.۲۶۷۵°	شانه
۰.۱۸۳۹°	۰.۲۱۳۹°	۰.۱۳۵۰°	آرنج
۰.۲۱۹۱°	۰.۲۷۴۵°	۰.۰۳۹۵°	مج

جدول ۱۶: خطای حالت ماندگار زوایای مفاصل دست مربوط به کنترل-

کننده RLPID در اثر تغییر پارامتر

[-5,-5,-5]	[5,5,5]	[1,1,1]	
۰.۵۹۷۱°	۰.۷۷۸۱°	۰.۴۰۶۷°	شانه
۱.۰۷۲۴°	۱.۰۱۱۰°	۰.۹۹۰۳°	آرنج
۰.۷۴۱۷°	۰.۵۱۸۴°	۰.۱۵۵۵°	مج

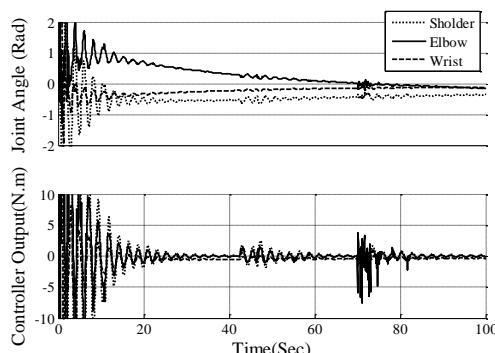
جدول ۱۴: خطای حالت ماندگار زوایای مفاصل دست مربوط به کنترل-

کننده RLPIDHTC در حضور اختشاش خارجی

[-5,-5,-5]	[5,5,5]	[1,1,1]	
۰.۵۲۳۵°	۰.۵۳۹۴°	۰.۲۴۷۱°	شانه
۰.۴۰۰۶°	۰.۴۱۵۳°	۰.۳۴۵۵°	آرنج
۰.۴۲۴۲°	۰.۴۹۱۸°	۰.۰۹۴۰°	مج

۴-۵- ارزیابی در حضور تغییر پارامترهای سیستم

در این مرحله، کارایی کنترل کننده‌های RLPID و RLPIDHTC در شرایطی که پارامترهای سیستم در طول زمان تغییر می‌کنند منظور بررسی قرار می‌گیرد. برای این منظور طول مفاصل، جرم لینک‌ها، طول و فاصله مرکز جرم تا مفصل را به اندازه $\pm 50\%$ مقدار اولیه آن تغییر می‌دهیم. نتایج مربوط به زوایه مفاصل و خروجی کنترلی کنترل کننده RLPID با در نظر گرفتن اثر تغییر پارامترها و شرایط اولیه [۵,۵,۵] در شکل (۱۷) نشان داده شده است. میانگین مربعات خطای زوایای مفاصل در شرایط اولیه مختلف نیز محاسبه شده و در جدول‌های (۱۵) و (۱۶) آمده است. عملکرد کنترل کننده RLPIDHTC نیز بازی در نظر گرفتن موقعیت اولیه [۵,۵,۵] برای مفاصل و اثر تغییر پارامتر، در شکل (۱۸) نشان داده شده است. میانگین مربعات خطای حالت ماندگار مربوط به زوایای مفاصل با شرایط اولیه مختلف در جدول‌های (۱۷) و (۱۸) آمده است.



شکل ۱۸: زوایای مفاصل و خروجی کنترل کننده RLPIDHTC برای سه مفصل شانه، آرنج و دست با شرایط اولیه [۵,۵,۵] در اثر تغییر پارامترهای سیستم.

جدول ۱۷: میانگین مربعات خطای زوایای مفاصل دست مربوط به کنترل کننده

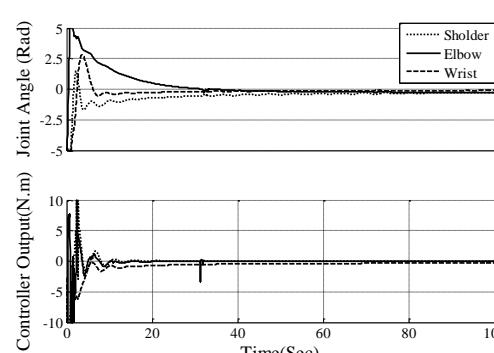
RLPIDHTC در اثر تغییر پارامتر

[-5,-5,-5]	[5,5,5]	[1,1,1]	
۰.۴۲۱۶°	۰.۴۸۶۹°	۰.۲۴۸۹°	شانه
۰.۲۹۸۸°	۰.۲۸۷۷°	۰.۲۳۵۶°	آرنج
۰.۲۲۶۴°	۰.۲۵۴۲°	۰.۰۴۱۰°	مج

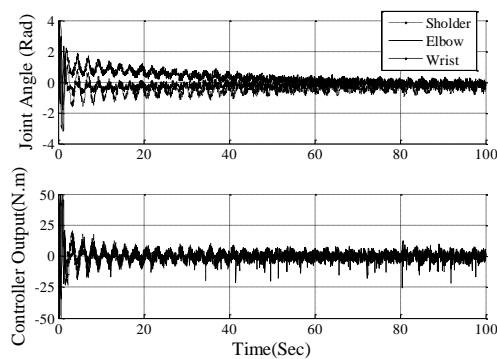
جدول ۱۸: خطای حالت ماندگار زوایای مفاصل دست مربوط به کنترل کننده

RLPIDHTC در اثر تغییر پارامتر

[-5,-5,-5]	[5,5,5]	[1,1,1]	
۰.۴۶۱۳°	۰.۵۲۹۷°	۰.۲۰۷۳°	شانه
۰.۴۶۴۰°	۰.۴۵۰۲°	۰.۳۸۶۷°	آرنج
۰.۴۴۰۹°	۰.۴۷۱۶°	۰.۰۹۵۷°	مج



شکل ۱۷: زوایای مفاصل و خروجی کنترل کننده RLPID برای سه مفصل شانه، آرنج و دست با شرایط اولیه [۵,۵,۵] در اثر تغییر پارامترهای سیستم.



شکل ۲۰: زوایای مفاصل و خروجی کنترل کننده RLPIDHTC برای سه مفصل شانه، آرنج و دست با شرایط اولیه $[5,5,5] = \theta_0$ در حضور نویز اندازه گیری.

جدول ۲۱: میانگین مربعات خطای زوایای مفاصل دست مربوط به کنترل کننده RLPIDHTC در حضور نویز اندازه گیری

$[-5,-5,-5]$	$[5,5,5]$	$[1,1,1]$	
0.3905°	0.4315°	0.3106°	شانه
0.3389°	0.3348°	0.2606°	آرنج
0.1810°	0.2686°	0.0724°	مج

جدول ۲۲: خطای حالت ماندگار زوایای مفاصل دست مربوط به کنترل کننده RLPIDHTC در حضور نویز اندازه گیری

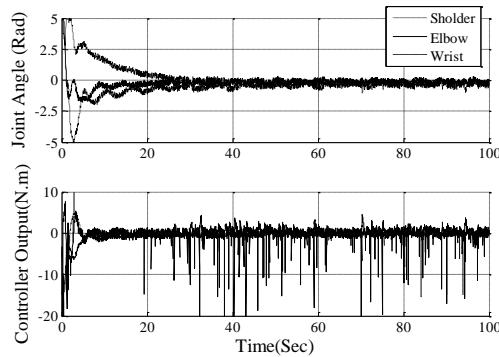
$[-5,-5,-5]$	$[5,5,5]$	$[1,1,1]$	
0.4284°	0.4747°	0.3226°	شانه
0.5055°	0.5008°	0.4189°	آرنج
0.3854°	0.4856°	0.2009°	مج

۵- ارزیابی مبتنی بر داده های انسانی

پس از شبیه سازی های کامپیوتری ملاحظه شد که کارایی کنترل کننده RLPIDHTC نسبت به RLPID بهتر است. لذا به منظور نوعی ارزیابی کاملتر و واقع گرایانه تر کنترل کننده RLPIDHTC، آزمایش های انسانی طراحی و انجام شدند. در آزمایش های انسانی الگوی فعالیت عضلات مفاصل دست، طی عملکرد حرکت دست با استفاده از سیگنال های الکترومایکروگرام عضلات، استخراج می شوند. چهار داوطلب سالم (۲۶-۲۶ ساله) راست دست در آزمایش ها شرکت داشتند، هیچ یک از افراد شرکت کننده دارای مشکلات حرکتی نبودند. به منظور ثبت سیگنال های عضلانی از یک دستگاه پاورلب ۴ کاناله و دو دستگاه با یو ام پ استفاده شد. نرم افزار استفاده شده به منظور ثبت لب چارت می باشد که قابلیت استفاده در محیط متلب را دارد. به منظور ثبت سیگنال های الکترومایکروگرام سطحی، تنظیمات زیر در سیستم قبل از شروع ثبت انجام شد. برای پوشش محدوده فرکانسی سیگنال EMG و نیز حذف انحراف از خط پایه و آریفکت های حرکتی و تداخلات مغناطیسی احتمالی، از یک فیلتر بالاگذر و یک فیلتر پایین گذر استفاده شد، همچنین برای حذف نویز برق شهر، فیلتر Notch اعمال گردید (فیلتر سیستم ثبت داده پاورلب). فرکانس نمونه برداری با در نظر گرفتن حداکثر فرکانس موجود در محتوای فرکانسی سیگنال EMG و

۶- ارزیابی در حضور نویز اندازه گیری

در این قسمت کارایی کنترل کننده های RLPID و RLPIDHTC در حضور نویز اندازه گیری به منظور بررسی می شود. به منظور لحاظ کردن عامل نویز در شبیه سازی ها، نویزی با توزیع یکنواخت با فاصله پیک تا پیک 0.5° به سیستم اعمال شده است. نتایج مربوط به زاویه مفاصل و خروجی کنترلی کنترل کننده RLPID با در نظر گرفتن اثر نویز اندازه گیری و شرایط اولیه [۵,۵,۵] در شکل (۱۹) آمده است. نتایج محاسبه میانگین مربعات خطای مفاصل و خطای حالت ماندگار زوایای مفاصل با شرایط اولیه مختلف در جداول های (۱۹) و (۲۰) آمده است. نتایج مربوط به زاویه مفاصل و خروجی کنترلی کنترل کننده RLPIDHTC نیز بازی شرایط اولیه [۵,۵,۵] و در نظر گرفتن اثر نویز اندازه گیری در شکل (۲۰) نشان داده شده است. میانگین مربعات خطای و خطای حالت ماندگار زوایای مفاصل نیز محاسبه شده و به ترتیب در جداول (۲۱) و (۲۲) آمده است.



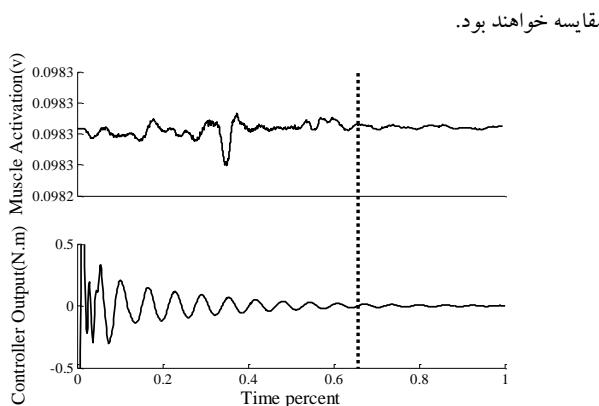
شکل ۱۹: زوایای مفاصل و خروجی کنترل کننده RLPID برای سه مفصل شانه، آرنج و دست با شرایط اولیه $[5,5,5] = \theta_0$ در حضور نویز اندازه گیری.

جدول ۱۹: میانگین مربعات خطای زوایای مفاصل دست مربوط به کنترل کننده RLPID در حضور نویز اندازه گیری

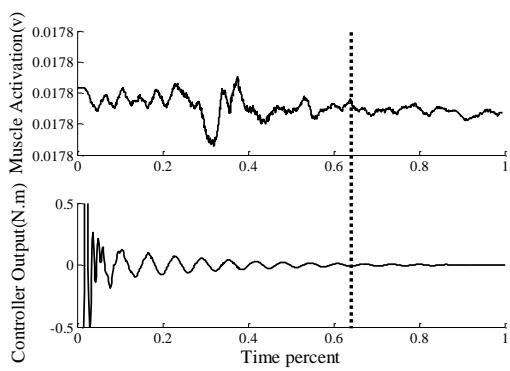
$[-5,-5,-5]$	$[5,5,5]$	$[1,1,1]$	
0.7999°	1.0326°	0.5188°	شانه
1.2988°	1.1552°	1.0667°	آرنج
0.4882°	0.3340°	0.0950°	مج

جدول ۲۰: خطای حالت ماندگار زوایای مفاصل دست مربوط به کنترل کننده RLPID در حضور نویز اندازه گیری

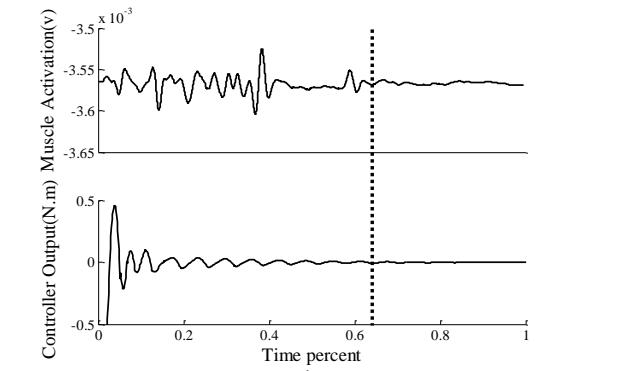
$[-5,-5,-5]$	$[5,5,5]$	$[1,1,1]$	
0.6991°	0.8496°	0.4600°	شانه
1.1098°	1.0422°	0.9992°	آرنج
0.6705°	0.5429°	0.2370°	مج



شکل ۲۲: خروجی کنترل کننده و الگوی فعالسازی عضله دلتوئید



شکل ۲۳: خروجی کنترل کننده و الگوی فعالسازی عضله ترایسپس براچی



شکل ۲۴: خروجی کنترل کننده و الگوی فعالسازی عضله اکستنسور کارپی ردیالیس برویس

۶- جمع‌بندی

در این پژوهش، یک راهکار کنترلی به منظور کنترل حرکت مدلی از بازوی دست با سه مفصل ارائه شد. در مدل مورد استفاده، اثر فعالسازی همزمان عضلات آنکونیست و آنتاگونیست مفصل مچ دست با استفاده از مکانیزم فریزسازی لحاظ گردید. با الهام از عملکرد سیستم اعصاب مرکزی در کسب مهارت‌های حرکتی، راهکار کنترلی ارائه شده مبتنی بر یکی از الگوریتم‌هاییادگیری تقویتی توسعه یافت. مطابق نتایج به دست آمده، سیستم کنترل حرکت بازوی دست مبتنی بر کنترل کننده RL نایاب‌دار است. به منظور بهبود عملکرد سیستم از یک کنترل کننده پس‌خورده، که این کنترل کننده PID بوده است، در کنار ساختار کنترلی RL استفاده شد. با توجه به نتایج مربوط به راهکار کنترلی RLPID، روشن می‌شود که کارایی کنترل کننده PID در کنار RL

اعمال نرخ نایکوئیست برای اطمینان از عدم تداخل برابر ۱ کیلوهertz و محدوده اندازه‌گیری دامنه سیگنال EMG، ۲ میلی‌ولت در نظر گرفته شد. در پایان نیز به منظور انجام پردازش پوش سیگنال گرفته شده است.

در این آزمایش الکترودهای ثبت بر روی عضلات دلتوئید^{۱۲} شانه، ترایسپس براچی^{۱۳} ساعد و اکستنسور کارپی ردیالیس برویس^{۱۴} قرار داده شده است. شکل (۲۱) سیستم آزمایش را نشان می‌دهد. از فرد خواسته شد در حالیکه نشسته است حرکات مسطوح را برطبق الگو رسم شده انجام دهد. در این آزمایشات سه حالت بررسی می‌شود، که بر طبق آن هر سه مفصل دست در زوایای ۶۰ درجه و ۴۰ درجه قرار می‌گیرند. به منظور هماهنگی ثبت-ها از سوژه می‌خواهیم ابتدا دست خود را در موقعیت موردنظر قرار دهد و منتظر علامت اول باشد، با شنیدن علامت اول سوژه دست خود را به سمت هدف، هر سه مفصل به سمت زاویه صفر درجه، حرکت می‌دهد. در پایان از سوژه درخواست شد تا زمانیکه علامت دوم را بشنود، دست خود را در موقعیت هدف نگه دارد.

انجام حرکات به سه درجه آزادی محدود شده و مفصل شانه مبدأ مختصات می‌باشد. هر آزمایش ۶ بار برای هر فرد تکرار و در کل برای هر فرد در مجموع سه حالت آزمایش، ۱۸ ثبت انجام گرفت. در بین هر حرکت به فرد فرست کوتاهی برای استراحت داده می‌شود تا از ایجاد خستگی عضلانی جلوگیری شود.



شکل ۲۱: نحوه انجام آزمایش‌های انسانی.

پس از ثبت داده‌های انسانی، زمانبندی تغییرات خروجی کنترل کننده در مطالعات شیوه‌سازی که همان گشتاور ایجاد شده در مفاصل هستند، با زمانبندی الگوی فعالیت عضلات که از داده‌های انسانی ثبت شده استخراج شده‌اند، مقایسه شده است. شایان ذکر است در مقایسه انجام شده تنها زمانبندی الگوی فعالیت عضلات و خروجی کنترل کننده مدنظر بوده است و هیچگونه مقایسه‌ای به لحاظ دامنه و فاز فعالیت عضلات و خروجی کنترل کننده انجام نشده است. در حقیقت تنها نوعی ارزیابی ابتدایی بر اساس داده‌های انسانی انجام گرفته است. شکل (۲۲) خروجی کنترل کننده پیشنهادی و الگوی فعالسازی عضله ترایسپس براچی در شکل (۲۳) نشان داده شده است. در شکل (۲۴) خروجی کنترل کننده پیشنهادی و الگوی فعالسازی عضله باز کننده مچ دست (کارپی ردیالیس برویس) آورده شده است. در واقع بوسیله نتایج ارائه شده در شکل‌های (۲۲) تا (۲۴)، زمانبندی فعالیت عضلات با زمانبندی فعالیت سیگنال‌های کنترلی قابل

¹²Deltoides¹³Triceps brachii¹⁴Extensor carpi radialisbrevis

می‌توان به منظور انجام فعالیت‌های روزانه زندگی برای افراد دچار فلج در بالا ته و پایین ته استفاده کرد [۲۳] - [۲۱]. افراد دچار آسیب‌های نخاعی اغلب برای حرکت دادن اندام‌های شان ناتوانند، اگرچه بیشتر اعصاب و عضلات آنها سالم هستند. تحریک الکتریکی عملکردی، این عضلات را برای انجام مجدد حرکات فعال می‌کند. کنترل حلقه بسته با استفاده از تحریک الکتریکی عملکردی، عملکرد عضله را بهبود می‌بخشد. در واقع آنچه در این مقاله ارائه شد، معرفی یک راهکار کنترلی حلقه بسته خودسازمانده مبتنی بر یادگیری تقویتی است که ضمن تضمین پایداری سیستم حلقه بسته و مقاوم بودن در برابر تغییر پارامترهای سیستم، اغتشاشهای خارجی و نویز اندازه‌گیری، قادر بوده است عملکرد کنترل حرکت مدلی از بازوی دست را با کارایی قابل قبولی انجام دهد. چنین راهکار کنترلی می‌تواند در راستای توسعه پروتزهای عصبی مربوط به کنترل حرکت دست، مثمر ثمر واقع شود.

مراجع

- [1] F.M.M.O. Camposand J.M.F. Calado, "Approaches to human arm movement control—A review," *Annual Reviews in Control*, vol. 33, Issue. 1, pp. 69-77, April. 2009.
- [2] JIANG Yu, JIANG Zhong-Ping and QIAN Ning, "Optimal Control Mechanisms in Human Arm Reaching Movements," *In Proc. 30th Chinese Control. Conf.* July 22-24. 2011.
- [3] Magnus J. E. Richardson and Tamar Flash, "Comparing Smooth Arm Movements with the Two-Thirds Power Law and the Related Segmented-Control Hypothesis," *J.Neurosci*, vol. 22, no. 18, pp. 8201-8211, 2002.
- [4] Elizabeth B. Torresand David Zipser, "Simultaneous control of hand displacements and rotations in orientation-matching experiments," *J. Appl. Physiol.*, vol. 96, pp.1978-1987, 2004.
- [5] Yasuhiro Wada, Yuichi Kaneko, Eri Nakano, Rieko OsuandMitsuoKawato, "Quantitative examinations for multi joint arm trajectory planningUsing a robust calculation algorithm of the minimum commanded torque change trajectory," *J Neural Networks*, vol. 14, pp. 381-393, 2001.
- [6] Toshikazu Matsui, KoukiTakeshita and Takahisa Shibusawa "Effectiveness of Human Three-Joint Arm's Optimal Control Model Characterized by Hand- Joint's Freezing Mechanism in Reproducing Constrained Reaching Movement Characteristics," *In Proc. Int. Joint. Conf.* pp. 1206-1211, 2009.
- [7] Weiwei Li, Emanuel Todorov and Xiuchuan Pan, "Hierarchical Feedback and Learning for Multi-joint Arm Movement Control," *In Proc. 27th Annu. Conf. EMBS*, 2005.
- [8] Kathleen M.Jagodnik and Antonie J. vandenBogert "Optimization and evaluation of a proportional derivative controller for planar arm movement," *J. Biomech.*, vol. 43, pp. 1086-1091, 2010.
- [9] Qinmin Yang and SarangapaniJagannathan, "Reinforcement Learning Controller Design for Affine Nonlinear Discrete-Time Systems using Online Approximators," *IEEE Trans. Syst. Man, Cybern.* vol. 42, no. 2, pp.377-390, April. 2012.
- [10] H.Kambara, J.Kim, M.Sato and Y.Koike, "learning arm's posture control using reinforcement learning and feedback-error-learning," *In Proc. 26th Annu. Int. Conf. EMBS* , vol. 1, pp. 486 - 489, 2004.
- [11] Shogo Uchiyama, Masanao Obayashi, Takashi Kuremoto and Kunikazu Kobayashi, "Robust Reinforcement Learning Control System with Ho Tracking Performance Compensator," *In Proc. 11th Cont. Auto. Syst. Conf.* 2011.
- [12] Magnus J. E. Richardson and Tamar Flash, "Comparing Smooth Arm Movements with the Two-Thirds Power Law and the Related Segmented-Control Hypothesis," *J.Neurosci*, vol. 22, no. 18, pp. 8201-8211, 2002.
- [13] Daniele Caligore, Eugenio Guglielmelli, Anna M. Borghi, DomenicoParisi and GianlucaBaldassarre "A Reinforcement Learning Model of Reaching Integrating Kinematic and Dynamic

مطلوب است. میانگین مربعات خطأ و خطای حالت ماندگار روش RLPID در مقایسه با محدوده تغییرات حرکت زوایای مفاصل کمتر است. نتایج نشان می‌دهد که کنترل کننده RLPID به ازای سه شرط اولیه مختلف برای مفاصل دست توانسته همگرایی با دقت قابل قبولی داشته باشد، البته نوساناتی دیده می‌شود که به دلیل دینامیک متغیر سیستم می‌باشد. با این حال به لحاظ تئوری اثباتی برای تضمین پایداری نداریم. به منظور تقویت این روش و تضمین پایداری سیستم از یک کنترل کننده HTC در کنار RLPID استفاده شد. از مقایسه روش‌های ارائه شده در قسمت نتایج روش می‌شود که بکارگیری هر دو کنترل کننده PID و HTC در کنار کنترل کننده RL بمنظور کنترل مدل سه مفصل دست با در نظر گرفتن مکانیزم فریزسازی قابل قبول است. همانطور که از جداول (۳) تا (۲۲) قابل مشاهده است، میزان میانگین مربعات خطأ و خطای حالت ماندگار RLPIDHTC در مقایسه با محدوده تغییرات حرکت دست و هچنین روش RLPIDHTC به میزان قابل توجهی کمتر است. در مجموع کنترل کننده RLPID به لحاظ حداقل فراجهش و عملکرد بهتری را به نسبت کنترل کننده RLPID به لحاظ حداقل فراجهش و طمان صعود داشته است. از طرفی این کنترل کننده در سه تست شبیه‌سازی طراحی شده به منظور ارزیابی ساختار کنترلی شامل اثر اغتشاش خارجی، تغییر پارامترهای سیستم و اثر نویز اندازه‌گیری نیز موفق عمل کرده است. میزان خطای محاسبه شده در سه تست شبیه‌سازی برای روش RLPIDHTC از روش RLPIDHTC به میزان قابل ملاحظه‌ای کمتر است. لذا ساختار کنترلی مبتنی بر RLPIDHTC روش پیشنهادی ما است.

در قسمت ارزیابی انسانی، الگوی فعالیت عضلات مفاصل دست، طی عملکرد حرکت دست با استفاده از سیگنال‌های الکترومایوگرام عضلات، استخراج شدند. همانطور که در شکل‌های (۲۲) تا (۲۴) نیز قابل مشاهده است شباهت قابل ملاحظه‌ای، به لحاظ زمانبندی تغییرات، بین خروجی کنترل کننده روشن پیشنهادی و میزان فعالیت عضلات وجود دارد. گشتاور ایجاد شده باعث تغییر فعالیت عضلانی می‌شود که کنترل کننده پیشنهادی به خوبی توانسته این تطابق با حالت طبیعی را ایجاد نماید. مجدداً تاکید می‌شود که در مقایسه الگوی فعالیت عضلات و خروجی کنترل کننده، تنها زمانبندی فعالیت و عدم فعالیت مدنظر بوده است و هیچگونه مقایسه‌ای به لحاظ دامنه و فاز فعالیت عضلات و خروجی کنترل کننده انجام نشده است. در واقع پس از اینکه دست به هدف رسیده و میزان تغییر فعالیت عضلات کاهش می‌یابد، خروجی کنترل کننده نیز نوسانات کمی دارد، به عبارتی گشتاور تولید شده کاهش می‌یافته است. نتایج نشان می‌دهند که همزمانی قابل ملاحظه‌ای بین زمان تغییرات سیگنال خروجی کنترل کننده با تغییرات الگو فعالیت عضلات دست فرد سالم وجود دارد. به عبارت دیگر نتایج نشان داده شده که به لحاظ بازه زمانی مربوط به فعالیت و عدم فعالیت عضلات و سیگنال خروجی کنترل کننده هم خوانی وجود دارد. البته به لحاظ دامنه تفاوت وجود دارد که دلیل آن عدم وجود عضله در مدل بیمار مجازی است و از طرفی مشخصات مدل بازو ارائه شده با حالت طبیعی یکسان نیستند.

یکی از موفق ترین راهکارهای موجود به منظور توانبخشی حرکتی، استفاده از پروتزهای عصبی است که از تحریک الکتریکی عملکردی به منظور فعال‌سازی عضلات از کار افتاده استفاده می‌کنند. در واقع از پروتزهای سیستم عصبی

- [18] MinijaTamosiunaite, TamimAsfour and FlorentinWörgötter, "Learning to reach by reinforcement learning using a receptive field based function approximation approach with continuous actions," *Biol. Cybern.* vol. 100, no. 1, pp. 249–260, 2009.
- [19] Kuo-Hsiang Cheng, "Auto-structuring fuzzy neuralsystem for intelligent control," *Journal of The FranklinInstitute*, vol.346, pp.267-288, 2009.
- [20] Sutton, R. S. and Barto, A. G., "Reinforcement learning: An introduction."
- [21] C.L.LYNCH and M.R.POPOVIC, "Functional Electrical Stimulation," *IEEE CONTROL SYSTEMS MAGAZINE*, pp. 40-50, april 2008.
- [22] P.S.Thomas, M.S.Branicky, A.V.D. Bogert, and K. Jagodnik, "FES Control of a Human Arm Using Reinforcement Learning" , 2007.
- [23] E.B.Marsolais and R.Kobetic, "Development of a practical electrical stimulation system for restoring gait in the paralyzed patient," *Clin. Ortho. & Rel. Res.*, vol. 233, pp. 64–74, 1988.
- [14] Philip Thomas, Michael Branicky,Antomie van den Bogert and Kathleen Jagodnik, "Application of the Actor-Critic Architecture to Functional Electrical Stimulation Control of a Human Arm," *In Proc. Innov. Appl. Artif. Intell.*, pp. 165–172, 2009.
- [15] Qinmin Yang, Jonathan Blake VanceandS.Jagannathan, "Control of Nonaffine Nonlinear Discrete-Time Systems Using Reinforcement-Learning-Based Linearly Parameterized Neural Networks," *IEEE Trans. Syst. Man, Cybern.* vol. 38, no. 4, pp.994-1001, August. 2008.
- [16] Lena AbbasiBrujeni, Jong Min LeeandSirish L. Shah, "Dynamic Tuning of PI-Controllers based on Model-free Reinforcement Learning Methods," *In Proc. Cont. Auto. Syst. Conf*, pp. 453-458, 2010.
- [17] F.L. LewisandKyriakos G. Vamvoudakis, "Optimal Adaptive Control for Unknown Systems Using Output Feedback by Reinforcement Learning Methods," *In Proc. 8th Cont. Auto. Conf*, 2010.