

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

زهرا حسن‌زاده بنابیدی<sup>۱</sup>، حمیدرضا کبروی<sup>۲</sup>، سعید طوسی‌زاده<sup>۳</sup>، رضا بوستانی<sup>۴</sup>

<sup>۱</sup> دانشجو کارشناسی ارشد مهندسی پزشکی، گروه مهندسی پزشکی، دانشگاه آزاد اسلامی، واحد مشهد، hasanzadeh\_67511@yahoo.com

<sup>۲</sup> استادیار، دانشکده فنی مهندسی، گروه مهندسی پزشکی، دانشگاه آزاد اسلامی، واحد مشهد، hkobravi@mshdiau.ac.ir

<sup>۳</sup> استادیار، دانشکده فنی مهندسی، گروه مهندسی برق، دانشگاه آزاد اسلامی، واحد مشهد، saeedtoosizadeh@yahoo.com

<sup>۴</sup> 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 میچ دست با استفاده از مکانیزم فریزسازی لحاظ 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 کنترل‌کننده کلاسیک مبتنی بر یادگیری تقویتی به منظور کنترل حرکت دست به سمت یک هدف مشخص، مورد ارزیابی قرار گرفت. نتایج نشان دادند که چنین کنترل‌کننده‌ای نمی‌تواند حتی بازای پس چندین اپیزود پایداری سیستم را حفظ نماید. در مرحله بعد به منظور بهبود وضعیت پایداری و همچنین کارایی کنترل‌کننده، ترکیبی از کنترل‌کننده تناسبی-انتگرالی-مشتقی (PID) و کنترل‌کننده مبتنی بر یادگیری تقویتی ارائه و مورد ارزیابی قرار گرفت. کنترل‌کننده 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 کنترل‌کننده HTC و کنترل‌کننده PID همراه با کنترل‌کننده مبتنی بر راهکار یادگیری تقویتی استفاده شده است. به این ترتیب پایداری سیستم کنترل حلقه بسته مبتنی تئوری لیاپانوف تضمین شده است. نتایج نشان دادند که کنترل‌کننده ترکیبی با استفاده از HTC در برابر اغتشاش خارجی و تغییرات تصادفی پارامترهای سیستم، کارایی بسیار خوبی داشته است. پس از شبیه‌سازی‌های کامپیوتری، به منظور ارزیابی کاملتر 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 الکترومیوگرام عضلات، استخراج شدند. نتایج نشان می‌دهند که همزمانی قابل ملاحظه‌ای بین بازه زمانی مربوط به فعالیت و عدم فعالیت تغییرات سیگنال خروجی کنترل‌کننده در مطالعات شبیه‌سازی با الگوی فعالیت عضلات دست فرد سالم وجود دارد.

**کلیدواژه:** کنترل‌کننده تناسبی-انتگرالی-مشتقی، کنترل‌کننده مقاوم (HTC)، مدل بازوی سه مفصل، مکانیزم فریزسازی، یادگیری تقویتی

## A Robust Control Strategy Based on Reinforcement Learning Approach to Rehabilitat 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

## ۱- مقدمه

آسیب نخاعی<sup>۱</sup> سطح بالا (سطح گردنی C1-C4)، می تواند موجب از دست دادن اکثر یا همه عملکردهای عضلانی ارادی در زیر گردن می باشد. تحریک الکتریکی عملکردی<sup>۲</sup> می تواند 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، گشتاورهای مکانیکی حول مفاصل، کنترل و به طبع آن حرکت ایجاد می شود. بدلیل دینامیک متغیر، اغتشاشات، رفتار متغیر با زمان و عدم قطعیت های موجود در سیستم های اسکلتی-عضلانی، توسعه سیستم های مبتنی بر FES نیازمند استفاده از روش های کنترلی هستند. یکی از زمینه های کاربردی تحریک الکتریکی عملکردی در ایجاد و کنترل حرکت بازو دست است. حرکت بازو یکی از رایج ترین فعالیت ها در زندگی روزمره ما است و چگونگی حرکات هماهنگ موضوع تحقیق های گسترده ای بوده است [۶]-[۱]. از رویکردهای کنترل حرکت بازوی انسان می توان به استفاده از مدل های توصیفی، مدل های کامل (شامل مدل های دینامیکی و مدل های تصادفی)، مدل های اجرایی حرکت و مدل های دینامیک مفاصل انسان اشاره کرد [۱]. هدف مدل های توصیفی، توصیف رفتار ظاهری حرکت انسان است. از جمله روابط تجربی این مدل ها که در شناسایی حرکات بازوی انسان مورد استفاده قرار می گیرد می توان به قانون فیت، پروفایل سرعت زنگوله وار در حرکات مستقیم و قانون ۲/۳ توان اشاره کرد. به رغم سادگی مدل های توصیفی، این مدل ها به علت در نظر نگرفتن جنبه های دینامیکی حرکت مورد انتقاد قرار می گیرند. از سوی دیگر مدل های دینامیکی قادر به پیش بینی برخی آزمایشات نبوده و در نتیجه رضایت بخش نیستند. نشان داده شده است هنگامیکه فردی دست خود را بین موقعیت اولیه و نهایی حرکت می دهد، ممکن است مسیر دست خود را به منظور اصلاح خط سیر، با استفاده از فیدبک های بصری تغییر دهد [۱]. مدل های کینماتیک این تطبیق پذیری<sup>۳</sup> را پیش بینی می کنند در حالی که مدل های دینامیکی قادر به انجام چنین کاری نیستند زیرا محدوده کاری مدل های دینامیکی در خط سیر تنها به موقعیت ابتدا و انتهای حرکت بستگی دارد و موقعیت میانی در نظر گرفته نمی شود [۱]. هاریس و ولپرت مفهوم متفاوتی از طراحی حرکت را بیان می کنند که ادغامی از مفاهیم دینامیک و کینماتیک می باشد [۱]. سجنوسکی و همکاران، [۱]، به برخی از جنبه های مهم نظریه هاریس و ولپرت تحت عنوان حداکثر دقت اشاره می کنند و بیان می دارند که مدل های مبتنی بر حداکثر بهره وری و حرکت بدون نوسان، نمی توانند توصیف خوبی از خود تطبیق پذیری سیستم عصبی در برداشته باشند. لی و همکارانش یک روش کلی برای کنترل پیش خورده سلسله مراتبی ارائه دادند که می تواند برای مشکلات کنترل حرکت دست مورد استفاده قرار گیرد [۷]. مدل دینامیکی ارائه شده توسط لی و همکارانش، یک

مدل دو سطحی، شامل مدل دینامیکی سطح بالا و مدل دینامیکی سطح پایین هستند که با یکدیگر هماهنگی دارند. در تحقیق دیگری جیانگ یو و همکارانش، تئوری کنترل ثابت بهینه پیوسته در زمان را برای مدل سازی سیستم های کنترل حرکتی انسان در حضور نویز وابسته به سیگنال، ارائه داده اند [۲]. مدل کنترلی آنها با برخی از مشاهدات مهم مانند عدم تقارن پروفایل سرعت سازگار است. جاجودنیک و همکاران سه کنترلر پس خورده تناسبی-مشتقی (PD) برای تحریک شش عضله بازو، با استفاده از دو سنسور زاویه مفصل ارائه داده اند [۸]. استفاده از راهکارهای کنترل ترکیبی مانند ترکیب کنترل پس-خوری و جلوسوی، یادگیری تقویتی و شبکه های عصبی مصنوعی نیز مورد ارزیابی قرار گرفته اند [۹]. اما با توجه به اینکه در فرآیند حرکت دست به سمت هدف، اولاً خط سیر مطلوبی برای تغییرات موقعیت دست قابل طراحی نیست و ثانیاً موقعیت نقاط ابتدا و انتها نیز ثابت نیستند، لذا استفاده از روش های کنترل مبتنی بر یادگیری تقویتی می تواند موثرتر باشد [۱۷]-[۱۰]. یادگیری تقویتی به معنای تصحیح عملکرد یک سیستم بر اساس پاداش و تنبیه های است که با توجه به شناخت از محیط و نتایج تعامل با محیط، طراحی شده اند. پژوهشگران دریافته اند که نرخ ارسال الکتریکی نورون های پخش کننده دوپامین در ناحیه تگمنتوم شکمی را می توان به تابع خطای الگوریتم تفاوت زمانی نسبت داد [۱۹]، که در فرایند حرکت دست به سمت هدف، این تابع در حقیقت وابسته به فاصله بین موقعیت فعلی و نهایی است. این امر موجب افزایش فعالیت مغزی در ناحیه گانگلیون بازال<sup>۴</sup> می گردد، که این بخش از ساختار مغز در کنترل حرکات عضله و یادگیری فرد دخالت دارد [۱۹]. با الهام از عملکرد سیستم اعصاب مرکزی در کسب مهارت های حرکتی، راهکار کنترلی مبتنی بر یکی از الگوریتم های یادگیری تقویتی توسعه داده شده است. به دلیل وجود عدم قطعیت، دینامیک های مدل نشده و عدم وجود مدل ریاضی مناسب در سیستم های اسکلتی-عضلانی، بکارگیری روش های یادگیری تقویتی در کنار کنترل کننده های مقاوم می تواند، راهکاری موثر باشد. بر این اساس هدف اصلی این تحقیق، ارائه راهکار کنترلی جدیدی بر مبنای ترکیب یک کنترلر با الگوریتم های یادگیری تقویتی می باشد. به منظور بهبود وضعیت پایداری و همچنین کارایی کنترل کننده، ترکیبی از یک کنترل کننده پس خورده 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 پس-خورده 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 لیاپانوف تضمین می شود. نتایج مطالعات مبتنی بر شبیه سازی های کامپیوتری نشان می دهند که راهکار کنترلی RLPIDHTC دارای بهترین کارایی بوده است. پس از ارزیابی بوسیله شبیه سازی کامپیوتری، به منظور ارزیابی کاملتر راهکار کنترلی پیشنهادی، آزمایش های انسانی طراحی و انجام شده اند. بطوریکه

<sup>۱</sup>Spinal Cord Injury(SCI)

<sup>۲</sup>Functional Electrical Stimulation (FES)

<sup>۳</sup>adaptation

<sup>۴</sup>Basal ganglia

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

۲-۱-۱- عملگر

عملگر یک کنترل کننده با عملکرد ردیابی بالا است، که دارای ساختار شبکه عصبی-فازی می باشد [۱۱]. سیگنال یادگیری تقویتی بصورت معادله (۱) بیان می شود، که از روش  $\epsilon$ -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 نرون های لایه میانی عملگر،  $w_j^{act}$  وزن از نرون لایه میانی  $\lambda$ ام به خروجی عملگر، n نوز برای یافتن سیگنال بهینه،  $b_j^{act}$  سازگاری نرون و  $\epsilon$  نرخ جستجو که در معادله (۲) آمده است.

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

که  $e_t$  خطای خروجی سیستم و  $\alpha_e$  ضریب ثابت است. سازگاری گره از حاصلضرب تابع عضویت گوسین بدست می آید، که بصورت معادله (۳) بیان می شود.

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

که c و  $\sigma$  مرکز و عرض تابع پایه لایه ورودی  $\lambda$ ام برای نرون  $\lambda$ ام میانی،  $e^i$  خطای خروجی سیستم مربوط به لایه ورودی  $\lambda$ ام می باشد و  $\bar{\omega}$  ثابت مثبتی است که از صفر شدن عرض تابع پایه جلوگیری می کند.

ساختار عملگر خودسازمانده است، به این صورت که تعداد نرون های لایه میانی شبکه عصبی بصورت خودکار اضافه و کم می شود تا ساختار ما بصورت موثر عمل نماید. مکانیزم خودسازمانده مشابه کار چنگ [۱۹] می باشد. افزودن نرون به این صورت انجام می گیرد که:

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

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

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

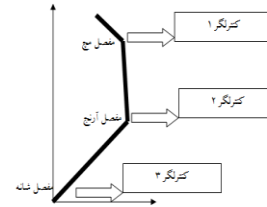
$$\Gamma_{\max}(t) \leq \Gamma_{th} \quad (5)$$

$\Gamma_{th} \in (0,1)$  آستانه افزودن به منظور اضافه کردن نرون است. اگر معادله (۵) برآورده شود، نرون جدید به لایه میانی اضافه می شود. هر نرون دارای یک شاخص اهمیت  $I_j$  است [۱۹] که نرون با استفاده از آن حذف می شود. به روزرسانی به اینصورت انجام می شود که:

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

۲- ارائه راهکار کنترل

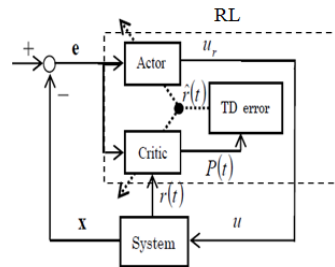
حرکت دست به سمت هدف، بوسیله ایجاد گشتاور حول سه مفصل مچ، آرنج و شانه ایجاد می شود. لذا برای کنترل چنین حرکتی دو راه حل وجود دارد. راه حل اول این است که، به از یک کنترل کننده چند ورودی-چند خروجی استفاده شود. اما طراحی این کنترل کننده مشکل بوده و به لحاظ پیاده سازی بار محاسباتی بالایی دارد، از طرفی تضمین پایداری چنین ساختاری بسیار مشکل است. راه حل دوم این است که یک استراتژی کنترل غیرمتمرکز 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-متمرکز، کوپلینگ های بین زیر سیستم ها، به عنوان دینامیک های مدل نشده یا اغتشاش برای هر زیر سیستم در نظر گرفته می شود.



شکل ۱: سیستم کنترل غیرمتمرکز به منظور کنترل حرکت بازوی دست ابتدا کنترل کننده مبتنی بر یادگیری تقویتی (RL) شامل دو قسمت 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؛ (۲) ارزیاب. عملگر سیگنال مطلوب را تولید کرده و ارزیاب وظیفه ارزیابی عملکرد شبکه عملگر را برعهده دارد، به عبارتی عملگر بر اساس خروجی های ارزیاب تطبیق می یابد [۱۸]. شکل (۲) ساختار کلی کنترل کننده مبتنی بر یادگیری تقویتی را نشان می دهد.



شکل ۲: ساختار کنترل کننده مبتنی بر یادگیری تقویتی

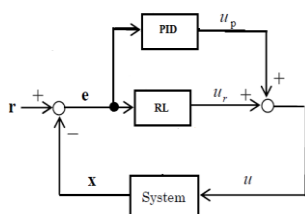
<sup>5</sup> Actor  
<sup>6</sup> Critic

## 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 پسخوردی به منظور کنترل حرکت دست و بهبود کارایی سیستم استفاده شده است. کنترل کننده تناسبی-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 و ارزیاب استفاده شده است. با توجه به اینکه مدل بازو دارای سه مفصل بوده است، برای کنترل هر مفصل از یک کنترل کننده PID با ساختار تشریح شده در شکل (۳) استفاده می‌شود. تابع تبدیل کنترل-کننده PID عبارتست از:

$$G_c(s) = K_p + K_I/s + K_Ds \quad (14)$$

که در آن  $K_p$  گین تناسبی،  $K_I$  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 PID و HTC

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  $H_\infty$  است، HTC<sup>۹</sup> (جبران‌ساز با دقت دنبال‌کنندگی 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  $H_\infty$  را تضمین نماید.

$$u_h = \frac{1}{8\tau^2} e^T P B, \quad B = [0 \ 0 \ \dots \ 0 \ 1]^T \quad (15)$$

$\tau$  ثابت است.

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

$\delta$  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 نیافتاده است، نقش موثرتری در کنترل سیستم برعهده دارد، اما پس از اینکه یادگیری به سطح

$\beta$  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،  $I_j$ ، برابر یک انتخاب شده است.

$$I_j \leq I_{th} \quad (V)$$

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 زمانی<sup>۷</sup> را به حداقل برساند. پاداش مورد نظر برابر است با:

$$P(t) = \sum_{j=1}^J w_j^{cri} b_j^{cri} \quad (8)$$

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

$$\hat{r}(t) = r(t) + \gamma \max P(t+1) - P(t) \quad (9)$$

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

$$r = \alpha_\theta e_t + \alpha_\theta e_t \quad (10)$$

$\alpha_\theta$  و  $\alpha_\theta$  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 خطا به منظور یادگیری شبکه عصبی استفاده می‌شود. در واقع یادگیری عملگر بر اساس معادله (۱۱) و یادگیری ارزیاب برطبق معادله (۱۲) خواهد بود.

$$\Delta w_j^{act} = \Pi_w^{act} \frac{\partial u}{\partial w_j^{act}} \hat{r} \quad (11)$$

$$\Delta w_j^{cri} = -\Pi_w^{cri} \frac{\partial P}{\partial w_j^{cri}} \frac{\partial}{\partial P} \left( \frac{1}{2} \hat{r}^2 \right) \quad (12)$$

$\Pi_w^{act}$  و  $\Pi_w^{cri}$  بترتیب نرخ یادگیری ارزیاب و عملگر و  $\hat{r}$  سیگنال خطا می‌باشند.

از روش پروجکشن<sup>۸</sup> برای محدود کردن  $w_j^{act}$  استفاده می‌کنیم، به این معنا که یک قانون به‌روزرسانی برای  $w_j^{act}$  بصورت زیر تعریف می‌کنیم:

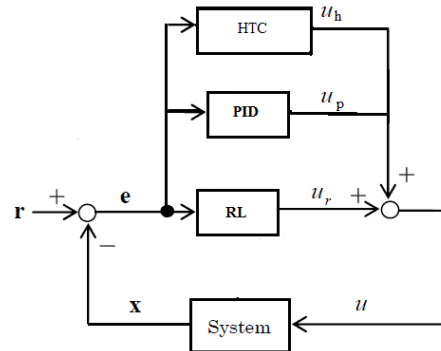
$$w_j^{act} = \begin{cases} w_j^{act} + \Delta w_j^{act} & \text{if } |w_j^{act} + \Delta w_j^{act}| < D \\ w_j^{act} & \text{otherwise} \end{cases} \quad (13)$$

که  $D$  محدوده موثر وزن می‌باشد. در واقع اگر قدرمطلق وزن پس از یادگیری از  $D$  کمتر باشد، به‌روز رسانی  $w_j^{act}$  انجام می‌گیرد.

<sup>۷</sup> Temporal Difference error (TD)

<sup>۸</sup> projection

مطلوبی رسید، هر دو کنترل کننده مبتنی بر یادگیری تقویتی و HTC در کنترل عملکرد حرکتی موثر خواهند بود. خروجی  $H_{\infty}$  با استفاده از حل P در معادله لیاپانوف ایجاد می شود.



شکل ۴: کنترل کننده مبتنی بر یادگیری تقویتی و کنترل کننده PID و  $H_{\infty}$  در روش RLPID اثبات تئوری برای تضمین پایداری سیستم کنترل وجود ندارد. اما در حضور کنترل کننده HTC می توان با تعریف تابع لیاپانوف پایداری سیستم کنترل حلقه بسته را اثبات کرد. در زیر بخش بعد پایداری سیستم اثبات می شود.

### ۲-۳-۱- آنالیز پایداری سیستم کنترل

سیستم غیر خطی مرتبه اول را بصورت زیر در نظر می گیریم:  
 $\dot{x} = A(x) + B(x)u$  (۱۶)  
 که x بردار حالت سیستم، f تابع پیوسته محدود ناشناخته، g تابع پیوسته مثبت محدود ناشناخته و u ورودی کنترل به سیستم می باشد.  
 تابع لیاپانوف را با در نظر گرفتن ماتریس قطری  $\Lambda$  که قطر اصلی آن شامل بهره پس خور k می باشد ( $\Lambda = k^T e$ ) که e خطای دنبال کنندگی، ماتریس مثبت معین P و ماتریس مثبت معین متعامد Q، که بصورت قراردادی تعیین شده، حل می نمایم با این شرط که پایداری هرویتز را برآورده نماید [۱۱].

$$\Lambda^T P + P \Lambda = -Q \quad (۱۷)$$

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

$$u^* = \frac{1}{B} (-A + \dot{r} + k^T e) \quad (۱۸)$$

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 می گیریم:

$$\dot{x} = A + Bu + Bu^* - Bu^* \quad (۱۹)$$

مشتق خطای ردیابی برابر است با:

$$\begin{aligned} \dot{e} &= \dot{r} - \dot{x} = (A + gu^* - k^T e) \\ &\quad - (A + Bu + Bu^* - Bu^*) \\ &= -k^T e + B(u^* - u) \end{aligned} \quad (۲۰)$$

معادله فضای حالت مشتق خطای ردیابی بصورت زیر می باشد:

$$\dot{e} = \Lambda e + B_1(u^* - u) \quad (۲۱)$$

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

$$u = u_r + u_h + u_p \quad (۲۲)$$

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

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

$$\dot{e} = \Lambda e + B_1(u^* - u_r - u_h - u_p) \quad (۲۳)$$

خطای تقریب سیگنال یادگیری تقویتی، کنترل کننده PID و ورودی بهینه به صورت زیر می باشد:

$$\varepsilon_u = u^* - u_r - u_p \quad (۲۴)$$

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

$$\dot{e} = \Lambda e + B_1(\varepsilon_u - u_h) \quad (۲۵)$$

برطبق معادله لیاپانوف، تطبیق به این صورت است:

$$V = e^T P e \quad (۲۶)$$

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

$$\dot{V} = \dot{e}^T P e + e^T P \dot{e} \quad (۲۷)$$

با استفاده از معادلات (۱۷) و (۲۳):

$$\dot{V} = -e^T Q e - 2e^T P B_1 u_h + 2e^T P B_1 \varepsilon_u \quad (۲۸)$$

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

$$\begin{aligned} \dot{V} &= -e^T Q e - \left\{ \left( \frac{\sqrt{g}}{2\tau} \right) e^T P B - 2\tau \left( \frac{g}{\sqrt{g}} \right) \varepsilon_u \right\} \\ &\quad + (2\tau \frac{g}{\sqrt{g}}) \varepsilon_u^T (2\tau \frac{g}{\sqrt{g}}) \varepsilon_u^2 \end{aligned} \quad (۲۹)$$

در نتیجه:

$$\dot{V} \leq -e^T Q e + (2\tau \frac{g}{\sqrt{g}}) \varepsilon_u^T (2\tau \frac{g}{\sqrt{g}}) \varepsilon_u \quad (۳۰)$$

از دو طرف معادله (۳۰) انتگرال می گیریم:

$$v(t_f) - v(0) \leq - \int_0^{t_f} e^T Q e dt + (2\tau \frac{g}{\sqrt{g}})^2 \int_0^{t_f} \varepsilon_u^T \varepsilon_u dt \quad (۳۱)$$

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

$$\int_0^{t_f} e^T Q e dt \leq e^T(0) P e(0) + (2\tau \frac{g}{\sqrt{g}})^2 \int_0^{t_f} \varepsilon_u^T \varepsilon_u dt \quad (۳۲)$$

که  $t_f$  زمان کنترل نهایی می باشد. همانطور که از معادله (۳۲) قابل مشاهده است، انتگرال خطای ردیابی e توسط مقادیر بدست آمده توسط ضریب خطای تقریب  $\varepsilon_u$  و ثابت میرایی  $\delta$  محدود شده است. به عبارتی اگر خطای تقریب و ثابت میرایی هر دو محدود باشد، خطای ردیابی نیز محدود است، در نتیجه پایداری سیستم کنترل اثبات می شود.

در نتیجه HTC بصورت معادله (۳۳) بیان می شود:

$$\int_0^{t_f} e^T Q e \leq e^T(0) P e(0) + \delta^2 \int_0^{t_f} \varepsilon_u^T \varepsilon_u dt \quad (۳۳)$$

که  $t_f$  زمان کنترل نهایی،  $\varepsilon_u$  مجموع عدم قطعیت خطای تقریب و اغتشاشات در زمان t و  $\delta$  ثابت مثبت غیر صفر می باشد [۱۱].

### ۳- مدل بازوی مجازی

مطالعات شبیه سازی مربوط به این تحقیق روی مدلی از بازوی دست انسان شامل سه مفصل انجام شده اند. شکل (۵) ساختار کلی این مدل را نشان می دهد [۶]. معادله دینامیکی بازوی سه مفصلی انسان که در صفحه افقی حرکت می کند، توسط معادله (۳۴) بیان می شود.

$$\begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix} \begin{bmatrix} \ddot{\theta}_1 \\ \ddot{\theta}_2 \\ \ddot{\theta}_3 \end{bmatrix} + \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} + \begin{bmatrix} D_1 & 0 & 0 \\ 0 & D_2 & 0 \\ 0 & 0 & D_3 \end{bmatrix} \begin{bmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \\ \dot{\theta}_3 \end{bmatrix} \quad (۳۴)$$

$$= M \ddot{\theta} + h + D \dot{\theta}$$

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

جدول ۱: پارامترهای دینامیکی بازو سه مفصل [۶]

پارامترها	شانه	آرنج	مچ
Li(m)	۰.۲۷۶	۰.۲۷۵	۰.۱۳۰
Lci(m)	۰.۱۴۱	۰.۰۸۶	۰.۰۴۳
mi(kg)	۱.۷۴۸	۰.۹۵۱	۰.۴۸۱
Ii(kg.m <sup>2</sup> )	۰.۰۱۲۷	۰.۰۰۴۱	۰.۰۰۰۵
Di(Nm.s/rad)	۰.۳۷۰	۰.۲۹۰	۰.۱۹۰

#### ۴- مطالعات شبیه‌سازی

مطالعات شبیه‌سازی مربوطه در محیط سیمولینک متلب 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 کنترل‌کننده‌ها در حضور اغتشاش، تغییر پارامترهای سیستم و نویز اندازه‌گیری مورد ارزیابی قرار گرفته‌اند. در ادامه نتایج مطالعات شبیه‌سازی به تفصیل ارائه خواهند شد.

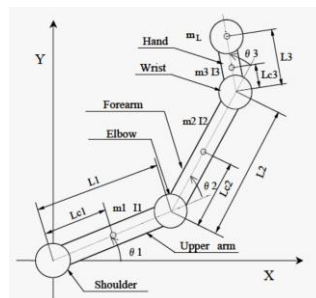
#### ۴-۱- کنترل‌کننده RL

همانطور که پیشتر عنوان شد، به منظور کنترل حرکت بازو از سه کنترل‌کننده مجزا برای هر مفصل استفاده می‌شود. در این قسمت سه کنترل‌کننده مبتنی بر یادگیری تقویتی (RL) به منظور کنترل حرکت مفاصل مورد استفاده قرار گرفته است. پارامترهای استفاده شده به منظور شبیه‌سازی کنترل‌کننده RL در جدول (۲) آمده است. کنترل‌کننده 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 [1,1,1]، [5,5,5] و [-5,-5,-5] در اپیزود اول نشان داده شده‌اند.

جدول ۲: پارامترهای استفاده شده در شبیه‌سازی سیستم کنترل‌کننده RL

k	{1,5}	$\gamma$	0.95
Q	diag{10,10}	$\alpha_\theta$	5
$\delta$	0.3	$\alpha_\theta$	1
g <sub>low</sub>	0.6	$\alpha_\varepsilon$	0.1
g <sub>up</sub>	1.5	$\eta_w^{crit}$	0.01
I	2	$\eta_w^{act}$	1
J	5	D	500
$\Gamma_{th}$	0.3	n	Random(-10-10)
$\sigma_c$	0.2	$\bar{\omega}$	0.001
I <sub>th</sub>	0.01		

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



شکل ۵: مدل بازوی ۳-مفصل انسان [۶]

لینک‌های ۱، ۲ و ۳ به ترتیب متناظر با بالای بازو، ساعد و مچ 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 لینک  $i$  و جرم جسم گرفته شده  $i$  ( $= 1.2 \text{ kg}$ ) هستند.  $L_i$  و  $L_{ci}$  به ترتیب طول، فاصله مرکز جرم تا مفصل نام ( $i=1,2,3$ ) 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 منقبض می‌شوند، افزایش می‌یابد [۶]. در این مکانیزم یک مولفه گشتاور مطابق رابطه (۳۵) به گشتاور مچ دست افزوده شده است، چراکه در اثر گرفتن جسم، بیشترین گشتاور حاصل از فعالیت عضلات آنتاگونیستی به مچ اعمال می‌شود. گشتاور پس‌خوردی حاصل از مکانیزم فریزسازی  $\tau_3^*$  در مفصل مچ دست بصورت زیر بیان می‌شود [۶]:

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

با لحاظ مکانیزم فریزسازی معادله

دینامیکی بازو به صورت معادله (۳۶) تغییر می‌یابد.

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

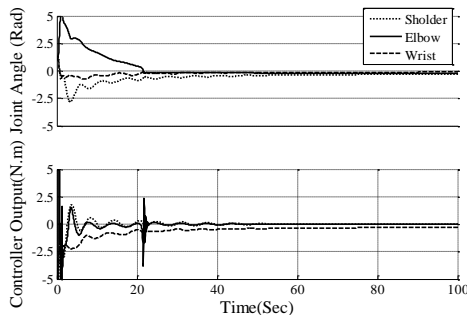
M ماتریس اینرسی، h بردار نیروی جانب مرکز و گریز از مرکز و D بردار نیروی ویسکوزیته می‌باشد. با این فرض که حرکت بازو در صفحه افقی و در ارتفاع شانه انجام می‌شود، معادلات حالت زیر بدست می‌آید:

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

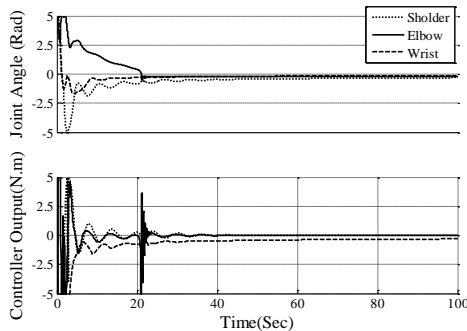
$$\theta = [\theta_1, \theta_2, \theta_3]^T, \tau = [\tau_1, \tau_2, \tau_3]^T, u = [\tau_1, \tau_2, \tau_3]^T$$

$$x = [\theta_1, \theta_2, \theta_3, \dot{\theta}_1, \dot{\theta}_2, \dot{\theta}_3, \tau_1, \tau_2, \tau_3]^T$$

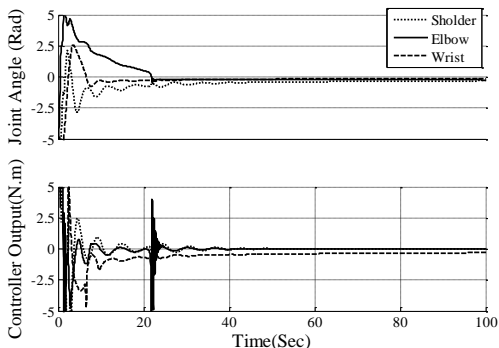
$k_{p2}$ ،  $k_{i1}$ ،  $k_{d1}$  و  $k_{p1}$  گین‌های مربوط به کنترل کننده مفصل شانه،  $k_{d3}$  و  $k_{i3}$ ،  $k_{p3}$  و  $k_{d2}$  و  $k_{i2}$  گین‌های مربوط به کنترل کننده مفصل آرنج و  $k_{p3}$ ،  $k_{i3}$  و  $k_{d3}$  گین‌های مربوط به کنترل کننده مفصل مچ دست می‌باشند، که ساختار کلی آن در شکل (۱) نشان داده شده است. میانگین مربعات خطا 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 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 (۶) محاسبه شده است.



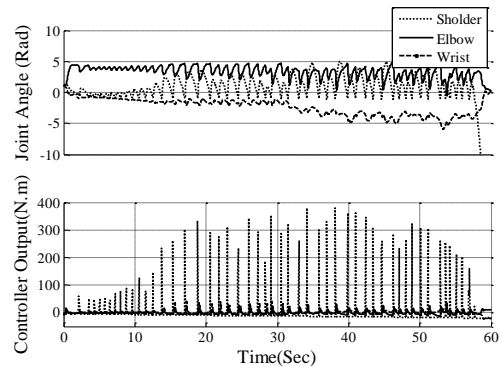
شکل ۹: زوایای مفاصل و خروجی کنترل کننده RLPID برای سه مفصل شانه، آرنج و دست با شرایط اولیه  $\theta_0 = [1,1,1]$



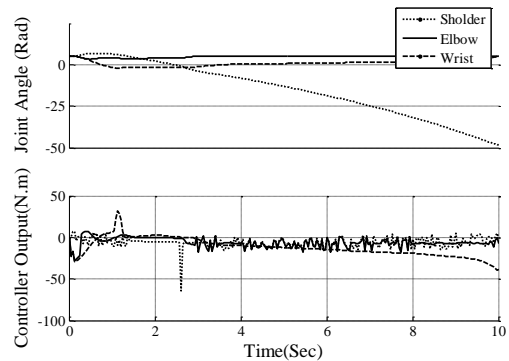
شکل ۱۰: زوایای مفاصل و خروجی کنترل کننده RLPID برای سه مفصل شانه، آرنج و دست با شرایط اولیه  $\theta_0 = [5,5,5]$



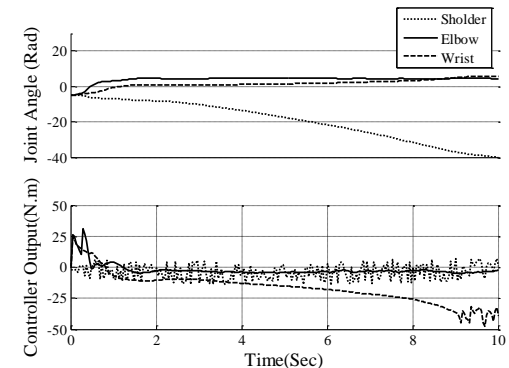
شکل ۱۱: زوایای مفاصل و خروجی کنترل کننده RLPID برای سه مفصل شانه، آرنج و دست با شرایط اولیه  $\theta_0 = [-5,-5,-5]$



شکل ۶: زوایای مفاصل و خروجی کنترل کننده RL برای سه مفصل شانه، آرنج و دست با شرایط اولیه  $\theta_0 = [1,1,1]$



شکل ۷: زوایای مفاصل و خروجی کنترل کننده RL برای سه مفصل شانه، آرنج و دست با شرایط اولیه  $\theta_0 = [5,5,5]$

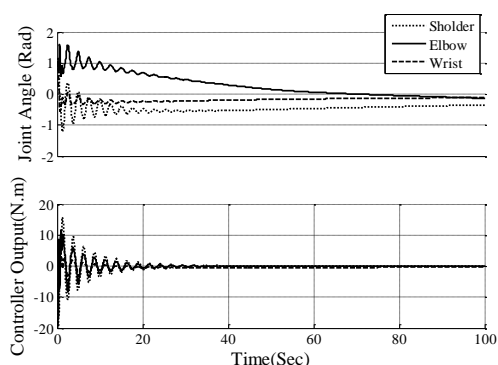


شکل ۸: زوایای مفاصل و خروجی کنترل کننده RL برای سه مفصل شانه، آرنج و دست با شرایط اولیه  $\theta_0 = [-5,-5,-5]$

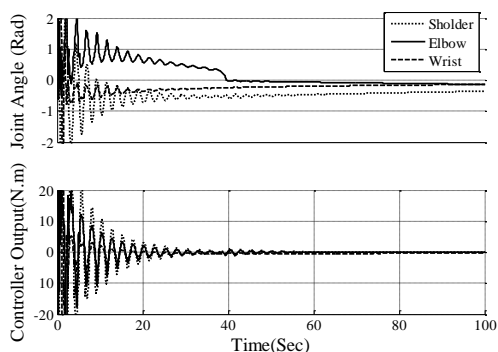
#### ۴-۲- کنترل کننده RLPID

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 کنار کنترل کننده RL استفاده خواهد شد. به منظور بررسی عملکرد کنترل کننده یادگیری تقویتی در کنار 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  $[1,1,1]$ ،  $[5,5,5]$  و  $[-5,-5,-5]$  به ترتیب در شکل‌های (۹) تا (۱۱) آمده است. پارامترهای کنترل کننده‌های PID بصورت زیر تنظیم شده‌اند:

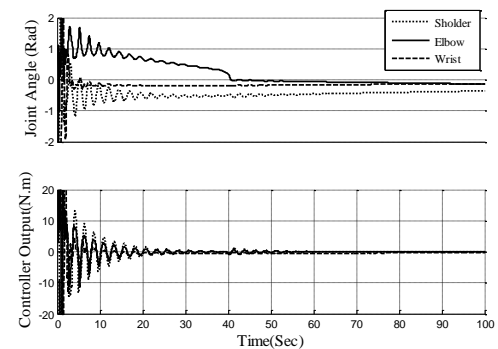
$$\begin{cases} K_{p1} = 2, K_{i1} = 0.5, K_{d1} = 1 \\ K_{p2} = 5, K_{i2} = 0.5, K_{d2} = 1 \\ K_{p3} = 2, K_{i3} = 1, K_{d3} = 1 \end{cases} \quad (38)$$



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 ۷: میانگین مربعات خطا زوایای مفاصل دست مربوط به کنترل کننده RLPIDHTC

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

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

	[1,1,1]	[5,5,5]	[-5,-5,-5]
شانه	۰.۴۸۹۶°	۱.۰۴۴۶°	۰.۶۷۴۸°
آرنج	۱.۰۵۲۷°	۱.۱۲۲۱°	۱.۱۸۵۶°
مچ	۰.۰۶۲۵°	۰.۳۰۰۲°	۰.۶۵۹۳°

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

	[1,1,1]	[5,5,5]	[-5,-5,-5]
شانه	۰.۴۲۵۵°	۰.۸۵۸۰°	۰.۶۰۵۲°
آرنج	۰.۹۹۱۵°	۱.۰۲۵۹°	۱.۰۵۶۴°
مچ	۰.۱۵۶۱°	۰.۵۱۲۰°	۰.۷۸۸۳°

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

	[1,1,1]	[5,5,5]	[-5,-5,-5]
شانه	-۲.۷۸۶۴°	-۵.۱۱۹۸°	۲.۱۱۱۱°
آرنج	-۰.۲۴۶۷°	-۰.۳۶۹۲°	۶.۰۵۹۶°
مچ	-۰.۷۴۸۴°	-۱.۶۷۹۱°	۲.۵۸۶۴°

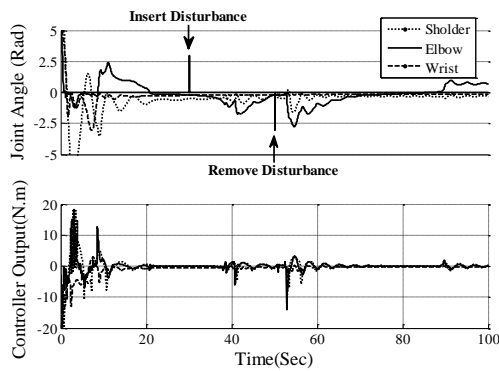
جدول ۶: زمان صعود (نزول) زوایای مفاصل دست مربوط به کنترل کننده

	[1,1,1]	[5,5,5]	[-5,-5,-5]
شانه	۵۲.۵۶	۱۴.۶۴	۱۵.۴
آرنج	۲۱.۸۴	۱۳.۶	۱۴.۵۲
مچ	۳۶.۵۲	۷.۱۲	۵.۶

### 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 معرفی 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 خطا 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.





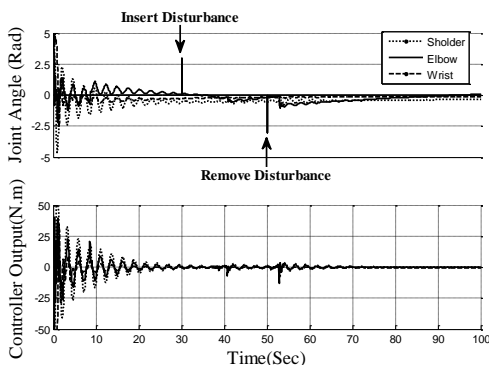
شکل ۱۵: زوایای مفاصل و خروجی کنترل کننده RLPID برای سه مفصل شانه، آرنج و دست با شرایط اولیه  $\theta_0 = [5,5,5]$  در حضور اغتشاش  $(td=30s,50s)$ .

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

	[-5,-5,-5]	[5,5,5]	[1,1,1]	
شانه	۰.۷۶۷۵°	۱.۵۸۰۳°	۰.۸۲۴۶°	
آرنج	۰.۸۰۵۲°	۰.۷۵۹۴°	۰.۶۷۹۸°	
مچ	۰.۳۰۸۰°	۰.۴۳۴۴°	۰.۰۹۲۵°	

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

	[-5,-5,-5]	[5,5,5]	[1,1,1]	
شانه	۰.۶۷۲۵°	۱.۱۲۱۷°	۰.۷۱۷۲°	
آرنج	۰.۸۸۳۲°	۰.۸۵۶۴°	۰.۸۱۱۱°	
مچ	۰.۵۱۲۲°	۰.۶۲۱۲°	۰.۲۳۵۰°	



شکل ۱۶: زوایای مفاصل و خروجی کنترل کننده RLPIDHTC برای سه مفصل شانه، آرنج و دست با شرایط اولیه  $\theta_0 = [5,5,5]$  در حضور اغتشاش  $(td=30s,50s)$ .

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

	[-5,-5,-5]	[5,5,5]	[1,1,1]	
شانه	۰.۴۱۱۶°	۰.۵۵۹۴°	۰.۱۸۴۹°	
آرنج	۰.۴۸۱۲°	۰.۴۴۵۴°	۰.۳۷۵۶°	
مچ	۰.۴۷۲۸°	۰.۴۸۱۳°	۰.۰۹۱۰°	

جدول ۹: حداکثر فرجهش و فروجهش زوایای مفاصل دست مربوط به کنترل کننده RLPIDHTC

	[-5,-5,-5]	[5,5,5]	[1,1,1]	
شانه	۴.۱۶۸۲°	-۴.۶۷۸۹°	-۱.۲۲۵۷°	
آرنج	۳.۴۰۱۵°	-۰.۶۰۰۴°	-۰.۱۳۷۵°	
مچ	۲.۹۱۱۶°	-۲.۵۴۱۳°	-۰.۵۵۹۵°	

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

	[-5,-5,-5]	[5,5,5]	[1,1,1]	
شانه	۴.۴	۱۰.۷۲	۸۱.۲۴	
آرنج	۱۲.۲۸	۱۴.۲۸	۴۶.۴۴	
مچ	۲.۱۶	۱.۸۴	۳۸.۰۸	

#### ۴-۴- ارزیابی در حضور اغتشاش خارجی

پس از ارزیابی‌های اولیه ملاحظه شد که کنترل کننده‌های غیرمتمرکز مبتنی بر RLPID و RLPIDHTC در کنترل موقعیت دست، کارایی قابل قبولی داشته‌اند. در ادامه به منظور مقایسه دقیق‌تر کارایی دو کنترل کننده، قابلیت‌آنها در مقابله با اعمال اغتشاش مکانیکی خارجی مورد بررسی قرار گرفت. به این منظور، در ثانیه ۳۰ پالس با دامنه ۳ Nm و در ثانیه ۵۰ پالس با دامنه ۳ Nm در مدت ۱ ثانیه به مفاصل دست اعمال شدند. عملکرد کنترل کننده RLPID بازای در نظر گرفتن موقعیت اولیه [5,5,5] برای مفاصل، در شکل (۱۵) نشان داده شده است. میانگین مربعات خطا و خطای حالت ماندگار مربوط به زوایای مفاصل با شرایط اولیه مختلف در جدول‌های (۱۱) و (۱۲) آمده است. عملکرد کنترل کننده RLPIDHTC نیز بازای در نظر گرفتن موقعیت اولیه [5,5,5] برای مفاصل، در شکل (۱۶) نشان داده شده است. میانگین مربعات خطا و خطای حالت ماندگار مربوط به زوایای مفاصل با شرایط اولیه مختلف در جدول‌های (۱۳) و (۱۴) آمده است.

جدول ۱۳: میانگین مربعات خطا زوایای مفاصل دست مربوط به کنترل-کننده RLPIDHTC در حضور اغتشاش خارجی

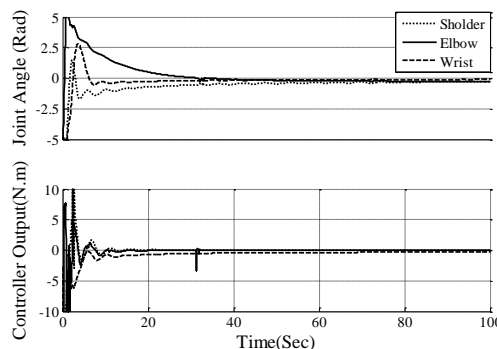
	[1,1,1]	[5,5,5]	[-5,-5,-5]
شانه	۰.۲۶۷۵°	۰.۵۳۰۴°	۰.۴۸۳۸°
آرنج	۰.۱۳۵۰°	۰.۲۱۳۹°	۰.۱۸۳۹°
مچ	۰.۰۳۹۵°	۰.۲۷۴۵°	۰.۲۱۹۱°

جدول ۱۴: خطای حالت ماندگار زوایای مفاصل دست مربوط به کنترل-کننده RLPIDHTC در حضور اغتشاش خارجی

	[1,1,1]	[5,5,5]	[-5,-5,-5]
شانه	۰.۲۴۷۱°	۰.۵۳۹۴°	۰.۵۲۳۵°
آرنج	۰.۳۴۵۵°	۰.۴۱۵۳°	۰.۴۰۰۶°
مچ	۰.۰۹۶۰°	۰.۴۹۱۸°	۰.۴۲۴۲°

#### ۴-۵- ارزیابی در حضور تغییر پارامترهای سیستم

در این مرحله، کارایی کنترلیکننده‌های 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  $\pm 5\%$  مقدار اولیه آن تغییر می‌دهیم. نتایج مربوط به زاویه مفاصل و خروجی کنترلی کنترلیکننده RLPID با در نظر گرفتن اثر تغییر پارامترها و شرایط اولیه [5,5,5] در شکل (۱۷) نشان داده شده است. میانگین مربعات خطا و خطای حالت ماندگار زوایای مفاصل در شرایط اولیه مختلف نیز محاسبه شده و در جدول‌های (۱۵) و (۱۶) آمده است. عملکرد کنترلیکننده RLPIDHTC نیز بازای در نظر گرفتن موقعیت اولیه [5,5,5] برای مفاصل و اثر تغییر پارامتر، در شکل (۱۸) نشان داده شده است. میانگین مربعات خطا و خطای حالت ماندگار مربوط به زوایای مفاصل با شرایط اولیه مختلف در جدول‌های (۱۷) و (۱۸) آمده است.



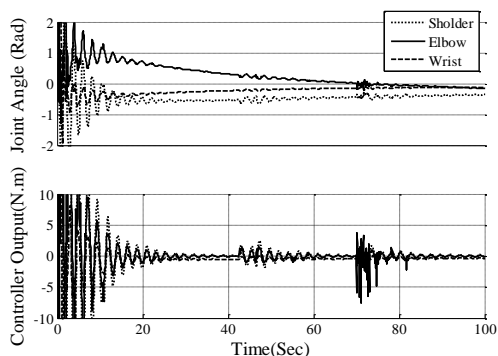
شکل ۱۷: زوایای مفاصل و خروجی کنترلیکننده RLPID برای سه مفصل شانه، آرنج و دست با شرایط اولیه  $\theta_0 = [5,5,5]$  در اثر تغییر پارامترهای سیستم.

جدول ۱۵: میانگین مربعات خطا زوایای مفاصل دست مربوط به کنترل-کننده RLPID در اثر تغییر پارامتر

	[1,1,1]	[5,5,5]	[-5,-5,-5]
شانه	۰.۴۷۴۱°	۰.۹۱۴۰°	۰.۶۶۵۳°
آرنج	۰.۱۰۵۰۲°	۰.۹۱۱۷°	۱.۲۱۹۶°
مچ	۰.۰۶۲۳°	۰.۳۰۶۸°	۰.۵۸۸۱°

جدول ۱۶: خطای حالت ماندگار زوایای مفاصل دست مربوط به کنترل-کننده RLPID در اثر تغییر پارامتر

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



شکل ۱۸: زوایای مفاصل و خروجی کنترلیکننده RLPIDHTC برای سه مفصل شانه، آرنج و دست با شرایط اولیه  $\theta_0 = [5,5,5]$  در اثر تغییر پارامترهای سیستم.

جدول ۱۷: میانگین مربعات خطا زوایای مفاصل دست مربوط به کنترلکننده RLPIDHTC در اثر تغییر پارامتر

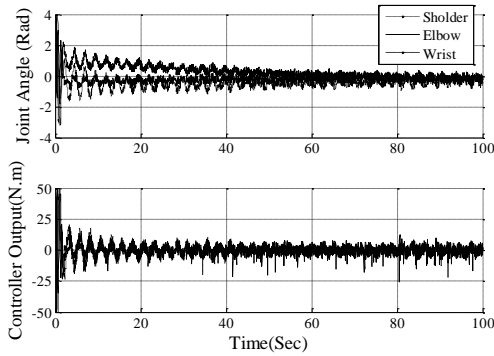
	[1,1,1]	[5,5,5]	[-5,-5,-5]
شانه	۰.۲۴۸۹°	۰.۴۸۶۹°	۰.۴۲۱۶°
آرنج	۰.۲۳۵۶°	۰.۲۸۷۷°	۰.۲۹۸۸°
مچ	۰.۰۴۱۰°	۰.۲۵۴۲°	۰.۲۲۶۴°

جدول ۱۸: خطای حالت ماندگار زوایای مفاصل دست مربوط به کنترلکننده RLPIDHTC در اثر تغییر پارامتر

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

۴-۶- 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 ۰.۵ 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 [5,5,5] در شکل (۱۹) آمده است. نتایج محاسبه میانگین مربعات خطا مفاصل و خطای حالت ماندگار زوایای مفاصل با شرایط اولیه مختلف در جدول های (۱۹) و (۲۰) آمده است. نتایج مربوط به زاویه مفاصل و خروجی کنترلی کنترل کننده RLPIDHTC نیز برای شرایط اولیه [5,5,5] و در نظر گرفتن اثر نویز اندازه گیری در شکل (۲۰) نشان داده شده است. میانگین مربعات خطا و خطای حالت ماندگار زوایای مفاصل نیز محاسبه شده و به ترتیب در جداول (۲۱) و (۲۲) آمده است.



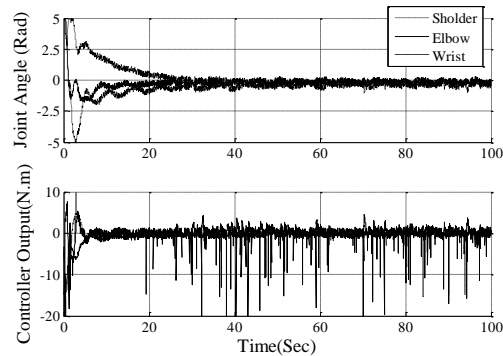
شکل ۲۰: زوایای مفاصل و خروجی کنترل کننده RLPIDHTC برای سه مفصل شانه، آرنج و دست با شرایط اولیه  $\theta_0 = [5,5,5]$  در حضور نویز اندازه گیری.

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

	[1,1,1]	[5,5,5]	[-5,-5,-5]
شانه	۰.۳۱۰۶°	۰.۴۳۱۵°	۰.۳۹۰۵°
آرنج	۰.۲۶۰۶°	۰.۳۳۴۸°	۰.۳۳۸۹°
مچ	۰.۰۷۲۴°	۰.۲۶۸۶°	۰.۱۸۱۰°

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

	[1,1,1]	[5,5,5]	[-5,-5,-5]
شانه	۰.۳۲۲۶°	۰.۴۷۴۷°	۰.۴۲۸۴°
آرنج	۰.۴۱۸۹°	۰.۵۰۰۸°	۰.۵۰۵۵°
مچ	۰.۲۰۰۹°	۰.۴۸۵۶°	۰.۳۸۵۴°



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

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

	[1,1,1]	[5,5,5]	[-5,-5,-5]
شانه	۰.۵۱۸۸°	۱.۰۳۳۶°	۰.۷۹۹۹°
آرنج	۱.۰۶۶۷°	۱.۱۵۵۲°	۱.۲۹۸۸°
مچ	۰.۰۹۵۰°	۰.۳۳۴۰°	۰.۴۸۸۲°

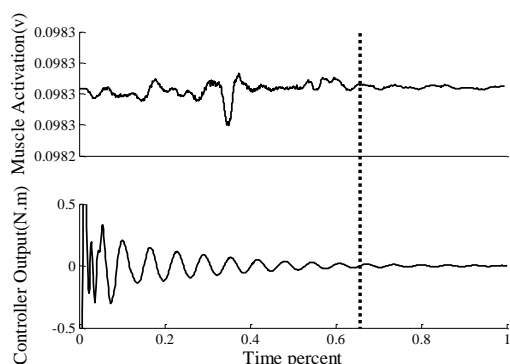
جدول ۲۰: خطای حالت ماندگار زوایای مفاصل دست مربوط به کنترل کننده RLPID در حضور نویز اندازه گیری

	[1,1,1]	[5,5,5]	[-5,-5,-5]
شانه	۰.۴۶۰۰°	۰.۸۴۹۶°	۰.۶۹۹۱°
آرنج	۰.۹۹۹۲°	۱.۰۴۳۲°	۱.۱۰۹۸°
مچ	۰.۲۳۷۰°	۰.۵۴۲۹°	۰.۶۷۰۵°

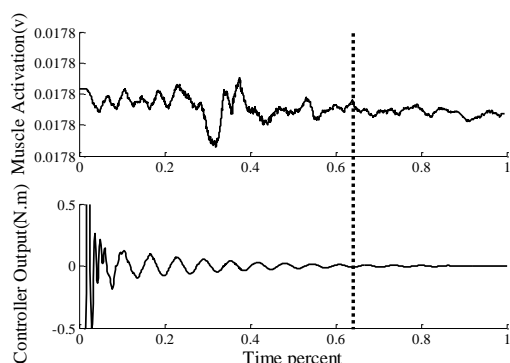
۵- ارزیابی مبتنی بر داده های انسانی

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 الکترومایوگرام عضلات، استخراج می شوند. چهار داوطلب سالم (۱۹-۲۶ 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 بایوآمپ 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 الکترومایوگرام 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 آرتیفکت 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

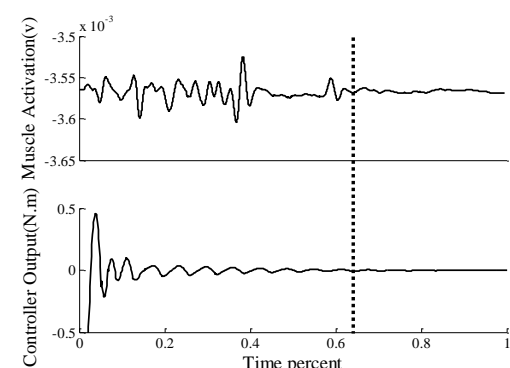
مقایسه خواهند بود.



شکل ۲۲: خروجی کنترل‌کننده و الگوی فعالسازی عضله دلتوئید



شکل ۲۳: خروجی کنترل‌کننده و الگوی فعالسازی عضله تریسپس براچی



شکل ۲۴: خروجی کنترل‌کننده و الگوی فعالسازی عضله اکستنسور کاری

ردیالیس برویس

## ۶- جمع‌بندی

در این پژوهش، یک راهکار کنترلی به‌منظور کنترل حرکت مدلی از بازوی دست با سه مفصل ارائه شد. در مدل مورد استفاده، اثر فعالسازی همزمان عضلات آگونیس و آنتاگونیست مفصل میچ 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 الگوریتم‌های یادگیری تقویتی توسعه یافت. مطابق نتایج به دست آمده، سیستم کنترل حرکت بازوی دست مبتنی بر کنترل‌کننده RL ناپایدار است. به منظور بهبود عملکرد سیستم از یک کنترل‌کننده پس‌خوردی، که این کنترل‌کننده PID بوده است، در کنار ساختار کنترلی RL استفاده شد. با توجه به نتایج مربوط به راهکار کنترلی RLPID، روشن می‌شود که کارایی کنترل‌کننده PID در کنار RL

اعمال نرخ نایکوئیست برای اطمینان از عدم تداخل برابر ۱ کیلوهرتز 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<sup>۱۲</sup> شانه، تریسپس براچی<sup>۱۳</sup> ساعد و اکستنسور کاری ریدیالیس برویس<sup>۱۴</sup> قرار داده شده است. شکل (۲۱) سیستم آزمایش را نشان می‌دهد. از فرد خواسته شد در حالیکه نشسته است حرکات مسطح 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، ۴۰ درجه و ۲۰ درجه قرار می‌گیرند. به منظور هماهنگی ثبت‌ها 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 مبدا 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 فرصت 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 کنترل‌کننده در مطالعات شبیه‌سازی 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 کنترل‌کننده مدنظر بوده است و هیچگونه مقایسه‌ای به لحاظ دامنه و فاز فعالیت عضلات و خروجی کنترل‌کننده انجام نشده است. در حقیقت تنها نوعی ارزیابی ابتدایی بر اساس داده‌های انسانی انجام گرفته است. شکل (۲۲) خروجی کنترل‌کننده پیشنهادی و الگوی فعالسازی عضله دلتوئید شانه را نشان می‌دهد. خروجی کنترل‌کننده پیشنهادی و الگوی فعالسازی عضله تریسپس براچی در شکل (۲۳) نشان داده شده است. در شکل (۲۴) خروجی کنترل‌کننده پیشنهادی و الگوی فعالسازی عضله بازکننده میچ دست ( کاری ریدیالیس برویس) آورده شده است. در واقع بوسیله نتایج ارائه شده در شکل‌های (۲۲) تا (۲۴)، نحوه زمانبندی فعالیت عضلات با زمانبندی فعالیت سیگنال‌های کنترلی قابل

<sup>12</sup>Deltoides  
<sup>13</sup>Triceps brachii  
<sup>14</sup>Extensor carpi radialisbrevis

می توان به منظور انجام فعالیت‌ها پروانه زندگی برای افراد دچار فلج در بالا تنه و پایین تنه 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، متمرکز واقع شود.

### مراجع

- [1] F.M.M.O. Campos and 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. 30<sup>th</sup> 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. Torres and 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 Osu and Mitsuo Kawato, "Quantitative examinations for multi joint arm trajectory planning Using a robust calculation algorithm of the minimum commanded torque change trajectory," *J Neural Networks*, vol. 14, pp. 381-393, 2001.
- [6] Toshikazu Matsui, Kouki Takeshita and Takahisa Shibusawa "Effectiveness of Human Three-Joint Arm's Optimal Control Model Characterized by Hand- Joint's Freezing Mechanism in Reproducing Constrained Movement Characteristics," *In Proc. Int. Joint. Conf*, pp. 1206-1211, 209.
- [7] Weiwei Li, Emanuel Todorov and Xiuchuan Pan, "Hierarchical Feedback and Learning for Multi-joint Arm Movement Control," *In Proc. 27<sup>th</sup> 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 Sarangapani Jagannathan, "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. 26<sup>th</sup> 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  $H_{\infty}$  Tracking Performance Compensator," *In Proc. 11<sup>th</sup> 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 Caligiore, Eugenio Guglielmelli, Anna M. Borghi, Domenico Parisi and Gianluca Baldassarre "A Reinforcement Learning Model of Reaching Integrating Kinematic and Dynamic

مطلوب است. میانگین مربعات خطا و خطای حالت ماندگار روش RLPID در مقایسه با محدوده تغییرات حرکت زوایای مفاصل کمتر است. نتایج نشان می‌دهد که کنترل‌کننده 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 کنترل‌کننده HTC در کنار RLPID استفاده شد. از مقایسه روش‌های ارائه شده در قسمت نتایج روشن می‌شود که بکارگیری هر دو کنترل‌کننده PID و HTC در کنار کنترل‌کننده 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 خطا و خطای حالت ماندگار روش RLPIDHTC در مقایسه با محدوده تغییرات حرکت دست و همچنین روش RLPID به میزان قابل توجهی کمتر است. در مجموع کنترل‌کننده RLPIDHTC عملکرد بهتری را به نسبت کنترل‌کننده RLPID به لحاظ حداکثر فراجش و طمان صعود داشته است. از طرفی این کنترل‌کننده 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 RLPID 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 کنترل‌کننده روش پیشنهادی و میزان فعالیت عضلات وجود دارد. گشتاور ایجاد شده باعث تغییر فعالیت عضلانی می‌شود که کنترل‌کننده پیشنهادی به خوبی توانسته این تطابق با حالت طبیعی را ایجاد نماید. مجدداً تأکید می‌شود که در مقایسه الگوی فعالیت عضلات و خروجی کنترل‌کننده، تنها زمانبندی فعالیت و عدم فعالیت مدنظر بوده است و هیچگونه مقایسه‌ای به لحاظ دامنه و فاز فعالیت عضلات و خروجی کنترل‌کننده انجام نشده است. در واقع پس از اینکه دست به هدف رسیده و میزان تغییر فعالیت عضلات کاهش می‌یابد، خروجی کنترل‌کننده نیز نوسانات کمی دارد، به عبارتی گشتاور تولید شده کاهش یافته است. نتایج نشان می‌دهند که همزمانی قابل ملاحظه‌ای بین زمان تغییرات سیگنال خروجی کنترل‌کننده 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 کنترل‌کننده هم‌خوانی 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 موفق‌ترین راهکارهای موجود به منظور توانبخشی حرکتی، استفاده از پروتزهای عصبی است که از تحریک الکتریکی عملکردی به منظور فعال-سازی 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 neural system 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.
- Control in a Simulated Arm Robot," *In Proc. 9<sup>th</sup> IEEE Int. Conf. development. learning.*, pp. 211-218, 2010.
- [14] Philip Thomas, Michael Branicky, Antonie 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. Conf.*, pp. 165–172, 2009.
- [15] Qinmin Yang, Jonathan Blake Vance and S. 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 Abbasi Brujeni, Jong Min Lee and Sirish 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. Lewis and Kyriakos G. Vamvoudakis, "Optimal Adaptive Control for Unknown Systems Using Output Feedback by Reinforcement Learning Methods," *In Proc. 8<sup>th</sup> Cont. Auto. Conf.* 2010.