A Multi-modal Feedback System for Ergonomic Pose Estimation

Kushal Thirani

Master Thesis
Mechanical Engineering
August 25, 2020

TU Delft Delft University of Technology

A Multi-modal Feedback System for Ergonomic Pose Estimation

by

Kushal Thirani

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Tuesday August,25 2020 at 10:30 am.

Student number:	4805569	
Project duration:	January,1 2020 – August,25 2020	
Thesis committee:	Prof. dr. ir. David Abbink,	TU Delft, chair
	Dr. Luka Peternel,	TU Delft, supervisor
	Prof. Dr.-Ing. Heike Vallery	TU Delft, external committee member

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft

Acknowledgements

I am immensely thankful to my supervisors Prof.dr.ir. David Abbink and Dr. Luka Peternel for providing me with valuable feedback and guiding me in the right direction throughout the duration of my thesis. I would also like to particularly thank them for carrying out long discussions and prompt meetings with me as and when needed.

I would in particular like to thank Dr. Luka Peternel, my daily supervisor who introduced me to the topic of ergonomics which in turn helped me shape it into a complete project.

I would also like to thank all my friends who took part in the human factors study carried out for this project and in turn made the overall study possible.

Last but not the least, I would also like to thank my family and friends at home for always being there for me and always supporting me in my endeavours.

Kushal Thirani
Delft, August 2020

CONTENTS

A Multi-modal Feedback System for Ergonomic Pose Estimation

Kushal Thirani (4805569)
Supervised by : Prof.dr.ir. David Abbink and Dr. Luka Peternel

Abstract— **This study proposes the creation of a multi-modal feedback system to guide humans towards ergonomic poses. A number of studies have tried to come up with methods where subjects are alerted upon crossing biomechanical or ergonomic thresholds while doing a task but not many have tried to successfully and efficiently guide users to ergonomic positions after having alerted them. Through this study we propose the creation of a multi-modal feedback system comprising of a visual and a speech based audio feedback and hypothesize that the proposed system will lead to a better performance as compared to the other feedback modalities when trying to guide users from one pose to another. During our study we have conducted two sets of experiments to carry out a comparative study between only audio, only visual and the proposed multi-modal feedback system to try and find the modality most effective and successful in guiding humans for pose corrections and a comparative study between two types of speech based audio feedbacks in joint space and end point space to motivate our choice for using the more desired one between the two for our proposed system.**

Speech based feedback in joint space came out as the preferred audio feedback due to its ability to allow users to carry out efficient and coordinated inter-joint movements especially in cases of high redundancy whereas the proposed multi-modal feedback system successfully shows its superiority over the other feedback modalities by showing equivalent results against the benchmark visual feedback when measured objectively and better results when measured subjectively due to its ability to successfully combine the advantages of audio and visual feedback and at the same time, avoid their limitations.

I. INTRODUCTION

Ever since the advent of the industrial revolution, the manufacturing sector has become central to the dependency of people on new products. This high demand for manufactured and fabricated goods has in turn led to an increase in involvement of people in the manufacturing process. The ever increasing demand for products with the increasing human population has in turn subjected people working in the manufacturing sector to both physical and mental stress. The continuous presence of physical stress which involves the application of force, repetitive execution of tasks or prolonged periods of unnatural postures has led to a dramatic increase in work related musculoskeletal disorders (WMSD). Humans are not designed to be working in factories or doing repetitive tasks. Hence, it is inherently unnatural to expect a human to work for prolonged periods under stressful conditions. According to a report published by the European Survey on Working Conditions (ESWC), ”24.7% of the European workers complain of backache, 22.8% of muscular pains, 45.5% report working in painful or tiring positions while 35% are required to handle heavy loads in their work” [11].

Multiple theories have been put forward which try to explain the cause of an injury. Kumar et al.,[2] had suggested that causes of injuries are either idiopathic or traumatic in nature. Feuerstein et al., [3] put forward a a theory suggesting that injuries are mostly caused because of a person’s work style. More such theories have been mentioned in literature [4],[12],[5] etc which suggest models and eventually try to establish causes of musculoskeletal injuries. But what is interesting to note is that all the theories suggest that the basic cause of an injury is biomechanical in nature.

Such work related musculoskeletal injuries can be prevented by alerting humans whenever they cross certain biomechanical thresholds (eg. joint torque overloads) while doing a task. There are numerous

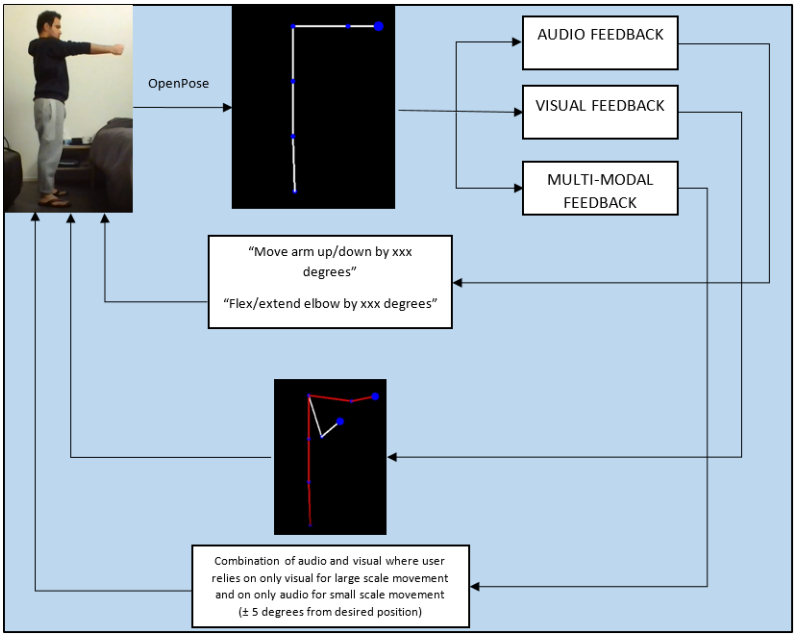


Fig. 1: The figure above shows the entire overview of the proposed study. The subject starts off from a starting pose which is captured by the camera and the image is used as input for a deep learning frame work (OpenPose[1]) used to detect the joint keypoints. OpenPose extracts the coordinates of the concerned joints and we use their positions to create a 5 DOF human body model as shown, to obtain the angular orientations and positions of the joints in real time. The current positions of the concerned joints are used as input for the different feedback methods. Each feedback method is then responsible for guiding the subject to efficiently and successfully carry out pose changes. Finally, a performance comparison is made between the three to try and prove our hypothesis.

studies which have successfully alerted humans through multiple feedback modalities such as visual [6], [7], audio[8], haptic and vibrotactile[9].

However, not many studies were found which used the above mentioned modalities to successfully and efficiently guide humans from unergonomic to ergonomic poses after having alerted them. The existing unimodal systems were found to have the following limitations :

Visual feedback systems have a major limitation of causing cognitive overloads to humans while doing a task as users have to constantly look elsewhere (eg. at a screen) to receive feedback regarding pose corrections and additionally, visual feedback could also be a problem for small error corrections as minute deviations might not be easily visible to the users.

Audio feedback systems on the other hand suffer from problems of habituation, where continuous presence of sound can very easily get lost in the background and the user realises this only when the audio feedback either stops or changes [15]. Further, transmission of complex multi-dimensional data for pose correction (name of joint, direction of movement and magnitude of movement) through sonification techniques (mapping of data into the different properties of sound such as frequency, amplitude, pitch etc.) or even through simple non-speech based audio feedback techniques could very easily overwhelm and confuse users.

Numerous haptic and vibrotactile sensors will have to be used in order to carry out a direction based guidance, the large number of sensors might hinder natural movement. Also, the perception of the magnitude of movement might differ from person to person given the subjective nature of the tactile feedback.

The above stated limitations of existing uni-modal systems motivated us to propose the creation of a system which would efficiently

and successfully guide humans from one posture to another if they were found to be crossing biomechanical thresholds of joint torque overloads while doing a task.

We believe that a multi-modal feedback system made up of a visual and speech based audio feedback would be best suited to guide humans from one pose to another. The visual feedback would very successfully help users with real time error visualization and position estimation and they could rely on it to very efficiently carry out large scale movements. Whereas the speech feedback could be used to very efficiently guide humans for minor pose adjustments once they are close enough to the desired position and no longer need to rely on visual feedback, thus also helping to redistribute the cognitive overload visual feedback systems suffer from. Additionally, the speech based feedback would also not suffer from the habituation problem simple audio feedback has and is known to be the most efficient type of audio feedback when providing users with instructions [15].

Based on the above findings we hypothesize that, **A multi-modal system made up visual and speech based audio feedback will lead to a better performance as compared to the other uni-modal feedback modalities when trying to guide users from one pose to another.** To try to test our hypothesis we carried out an experiment explained ahead.

II. METHOD

A. Feedback Mechanisms

In this section we explain the multiple feedback mechanisms created to carry out a comparative study of performance measures in order to test our hypothesis. In total, we created four types of feedbacks. Two types of speech based audio feedback methods, a visual feedback system and finally, the proposed multi-modal feedback system.

1) Audio Feedback in Joint Space: Speech based audio feedback in joint space involves a feedback method where the user is asked to manipulate his pose in a joint specific manner through speech based commands. The commands are limited to the shoulder and elbow joints for our case. The commands are in English and are output at a rate of 180 words per minute which is slightly higher than the average of 140-160 words per minute worldwide. The direction and magnitude of movement are relative to the current position of the joints.

Once the user has oriented his/her joints in the desired pose, the system alerts users by raising a beeping noise. The table below (See Table:I) shows the list of commands users receive for this particular modality.

TABLE I: List of speech based audio commands in Joint Space feedback

Command	Joint	Meaning
”Move arm up by xxx degrees”	Shoulder	Shoulder flexion
”Move arm down by xxx degrees”	Shoulder	Shoulder extension
”Flex elbow by xxx degrees”	Elbow	Elbow flexion
”Extend elbow by xxx degrees”	Elbow	Elbow extension

2) Audio Feedback in End point Space: Another speech based audio feedback mechanism was created which carried out arm manipulations in end point (Cartesian) space. The wrist joint was assumed to be the end point for our study. Thus, subjects no longer received joint specific commands but rather received commands which asked them to manipulate their end point to a desired coordinate position.

The commands were in English and are output at 180 words per minute, same as the previously mentioned audio feedback method. The direction and magnitude of movement was relative to the current position of the end point and the unit of movement was in centimeters. The table below (See Table:II) shows the list of commands subjects received for this particular modality.

TABLE II: List of speech based audio commands in End Point Space feedback

Command	Meaning
”Move arm back by xxx”	Move EP towards the body
”Move arm forward by xxx”	Move EP away from body
”Move arm down by xxx”	Move EP vertically down
”Move arm up by xxx”	Move EP vertically up

3) Visual Feedback: Visual feedback system was assumed to be the benchmark feedback modality when it comes to performance measurement. This is due to a number of reasons - firstly, concurrent visual feedback is the easiest and most natural form of feedback modality[10]. Secondly, visual feedback is superior to other senses when it comes to understanding spatial information[10] and finally, humans usually rely on visual feedback when it comes to following trajectories and interacting with the environment[16].

In visual feedback, subjects were able to see a real time, stick figure representation of themselves (in red) and the target position (in white). The blue circles represented the five joints of the human body namely ankle, knee, hip, shoulder and elbow and the larger blue circle at the end represents the wrist (See Fig: 2)

The subjects through this feedback were asked to manipulate their position and try to reach the desired pose in white and upon successfully doing so, the colour of the screen would change to green.

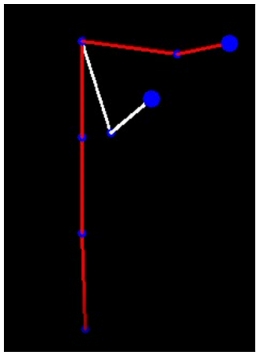


Fig. 2: The image above shows visual feedback which was provided to the subjects. The current pose of the subject is shown in real time in red and the desired pose is shown in white. Subjects are tasked to reach the desired pose and upon successfully doing so, the colour of the screen changes to green.

4) Multi-modal Feedback: The proposed multi-modal system interactively combines the two modalities of visual and the speech based audio feedback in joint space as mentioned before. The proposed system can be used by users to first carry out larger displacements by relying on the visual feedback modality and once they are close enough to the desired pose($\pm$ 5 degrees) they can then rely completely on audio feedback to carry out minor adjustments. This way users will not have to constantly look up at the screen to receive visual feedback regarding their pose and can instead concentrate on the task while receiving audio feedback for minor adjustments and only switch to visual feedback in case they displace their joints by a large amount. Upon reaching the desired pose, the user is alerted by receiving a beeping noise.

B. Participants

The study was carried out with 14 participants (10 male, 4 female). The participants were between 23 and 27 years old (mean = 25.14 years; std deviation = 0.989 years) with none/or corrected vision and/or hearing impairment.

C. Experimental Design

Two within subject experiments were conducted to try and prove the need for a multi-modal feedback system for ergonomic pose estimation by comparing user performance between only audio feedback, only visual feedback and the proposed multi-modal feedback system and another, to motivate our choice of using joint space audio feedback in the proposed multi-modal system. For each case, the performance results (obtained through the chosen metrics explained in the next section) for the audio and multi-modal feedback system were compared against the visual feedback system, which we have assumed to be a benchmark. We believe that by carrying out a performance comparison between the three feedback techniques we will be able to find the modality which was most effective and successful in carrying out guided pose correction and this in turn would help us test our previously mentioned hypothesis.

D. Dependent Variables

To judge user performance we decided to use both objective and subjective measures for each experiment.

1) Objective Measures:

- **TASK COMPLETION TIME -** Task completion time (TCT) measures the time taken to complete the task which is indicated by the start time till the time when a user holds the desired position for 10 seconds. A smaller TCT would indicate good performance as it would mean that the user was able to understand the feedback properly and could thus complete the task quicker (See Fig:3).

- **TOTAL DISTANCE MOVED -** Total distance Moved refers to the sum of the differences of every joint angular position for every time step which is measured until the user is successfully able to orient both his joints (shoulder and elbow) within the threshold ($\pm$ 5 degrees) of the desired angular position. A larger distance moved would be an indication of bad performance as it would show the user's inability to understand the feedback properly causing them to make multiple adjustments (due to repetitive overshooting) to their pose before reaching the desired position.(See Fig:3)

2) Subjective Measures:

- **NASA-TLX -** The NASA-TLX was one of the subjective measures used to evaluate the overall task load by calculating a weighted average of 6 different measures namely mental demand, physical demand, temporal demand, performance, effort and frustration [14]. A final task load score would be achieved by taking a weighted average of the above mentioned metrics and a low score would be an indication of low task load and would help us in judging the feedback mechanism from a subjective point of view.

- **VAN DER LAAN -** This subjective questionnaire was used to check the acceptance of the desired feedback modality by getting a usefulness and satisfaction score [13]. Each score ranges between -2 to +2 and finally a plot between the scores would help make a case for the acceptability of the feedback mechanism. A final score in the upper right quadrant would indicate user acceptability.

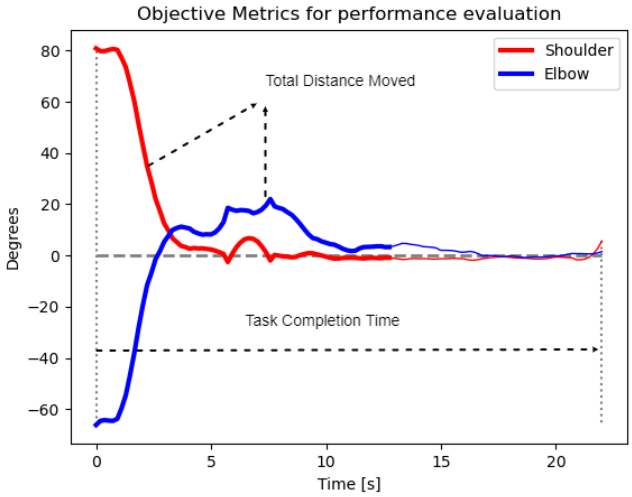


Fig. 3: The image above shows the filtered sensor data for the shoulder and elbow joint for a single task. The final pose angles have been subtracted from both joints so that they converge to zero. Task Completion Time (TCT) is a measure of the time taken to complete the task whereas Total Distance Moved (TDM) measures the sum of the difference between the angular positions between every time step. TDM is measured until the end of the highlighted line as beyond that the user is tasked to try and hold their position.

E. Procedure

For the experiment, the subjects underwent 5 trials of each of the three conditions. To familiarize the subjects to the feedback conditions each subject underwent two or more practice trials and once the subject was comfortable with the feedback, the experiment was started.

The subjects were asked to stand opposite to a camera system which was responsible for capturing their poses and joint positions. We used an open source pose estimation software, OpenPose [1] to capture and evaluate these joint angles (Refer to Appendix A for details). All angular measurements were done in the saggital plane for the right side only. The subjects were asked to stand in a neutral starting pose which was either standing with their arm stretched out (shoulder at 90 degrees and elbow at 0 degrees) or their arm stretched over their head (shoulder and elbow both at 0 degrees). These two positions were chosen because it was easiest to estimate

the starting angles of the joints with the naked eye. They were then asked to follow the feedback commands and try to reach the desired pose. Their main objective was to try and complete the task as quickly as possible but at the same time avoid unnecessary errors and thus achieve good performance. In order to minimize learning effect, different starting and end poses were selected for the 5 repetitions which were shuffled across different conditions.

After finishing the 5 trials for each condition, the subjects were asked to fill in the NASA-TLX and Van der Laan questionnaires.

III. RESULTS

A. Task Completion Time

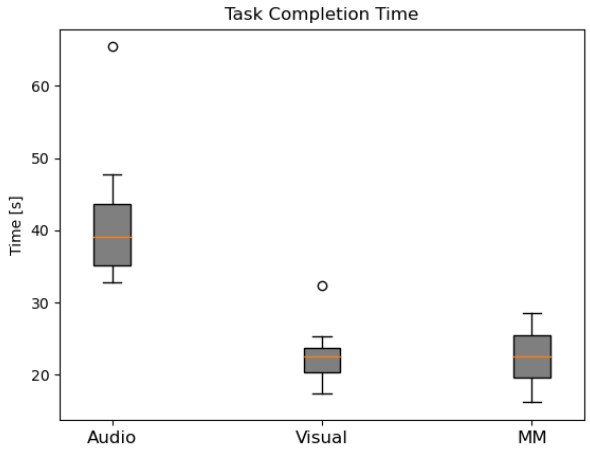


Fig. 4: Box plot of the total time to task completion for each condition can be seen above. The proposed multi-modal feedback system shows similar task completion time as compared to the benchmark visual feedback while audio feedback can be seen to have the highest TCT.

Audio feedback can be seen to have the highest task completion time (mean = 40.971 seconds, std deviation = 8.038 seconds) when compared to visual feedback (See Fig: 4)(mean = 22.485 seconds, std deviation = 3.615 seconds) ($p < 0.05$).
The multi-modal feedback system on the other hand can be seen to have a similar task completion time (mean = 22.285 seconds, std deviation = 3.857 seconds) when averaged for the 14 participants and compared to visual (p=0.89)(See Fig: 4).

B. Total Distance Moved

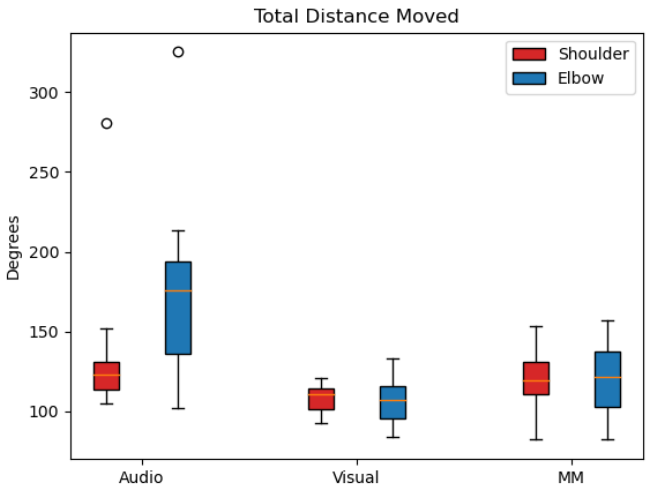


Fig. 5: A box plot showing the total distance moved by the individual joints during the task for every condition. Similar results can again be seen between the visual and multi-modal feedback systems.

Only audio feedback can be seen to have a higher overall movement for shoulder (mean = 134.533 degrees, std deviation = 42.462 degrees) and elbow (mean = 172.311 degrees, std deviation = 54.801 degrees) when compared to visual feedback for shoulder (mean = 108.396 degrees, std deviation = 7.854 degrees) and elbow (mean = 106.419 degrees, std deviation = 14.856 degrees) ($p<0.05$ for each joint)
Multi-modal feedback on the other hand can be seen to have a slightly larger overall movement for either joint (shoulder - mean = 119.627 degrees, std deviation = 16.339 || elbow - mean = 119.512 degrees, std deviation = 24.205 degrees) when compared to visual ($p<0.05$ for shoulder and p=0.1 for elbow). (See Fig: 5)

C. Subjective Measures

Audio feedback can be seen to have an average overall task load index of 42.162 and visual feedback can be seen to have an average overall task load index of 41.496 making both feed backs almost equally demanding with no significant

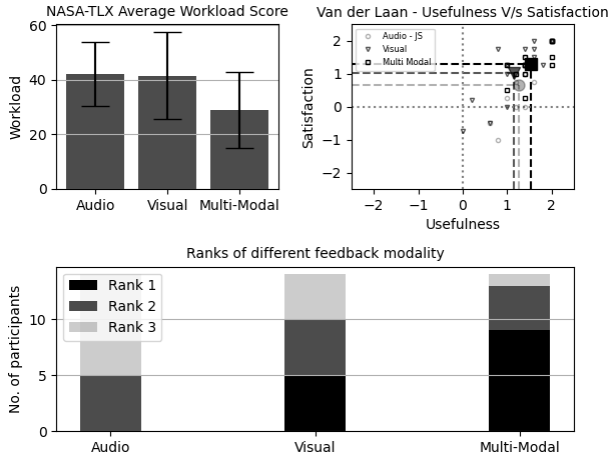


Fig. 6: The figure above shows the results for the three subjective measures used to evaluate task performance. The proposed multi-modal feedback system can be clearly seen to have better subjective experience compared to the other feedback modalities in terms of overall task load, acceptance and user preference.

difference (p=0.90). Multi-modal feedback shows an average overall task load index of 28.996 which is significantly lower as compared to visual feedback ($p<0.05$).

When comparing the results of the Van der Laan questionnaire to see which feedback system is deemed to be most acceptable by users, the multi-modal system stands out to be a clear favourite by having higher scores on both usefulness and satisfaction scales.

We also asked users to rank the systems between 1-3 (1 being the best and 3 being the worst) to further corroborate the results of the subjective measures and know their preference. The multi-modal feedback system stood out to be the clear favourite with 9 out of 14 people voting it to be their most preferred form of feedback for a guidance task. (See Fig:6)

IV. DISCUSSION

A. Task Completion Time

Audio based feedback can be seen to have the highest task completion time among the three, when averaged for the 14 participants. This was an expected result as subjects took more time because of two main reasons firstly, information transmission through audio in a speech based format is slow and thus subjects had to wait and listen to the entire command and only then could they carry out the movement. Secondly, users did not know how far or close they were to the desired position and the added difficulty of approximating angles mentally, led to a constant over or undershooting of the target position. These two reasons together were responsible for a large task completion time for audio feedback. Visual feedback on the other hand allowed users to be aware of their current positions at all times and at the same time made users aware of the difference between their current and desired pose. This way users did not overshoot their mark and were able to control their movements as soon as they got close to the desired pose.

The multi-modal feedback system on the other hand can be seen to have a similar TCT when averaged for 14 participants and compared to visual even though subjects had to switch from one modality to another. This again was an expected result since in the Multi-modal feedback, the subjects relied on the visual feedback to carry out larger movements and could thus visualize their position differences making them move quicker. The audio feedback was activated only when the subject was +- 5 degrees away from the desired position, thus limiting the comparatively slower audio feedback to only a small portion of the task. Given that we had assumed visual feedback to be a benchmark, a similar TCT does show that the new feedback modality is equally efficient for the current metric.

B. Total Distance Moved

Audio based feedback can be very clearly seen to have a significantly larger distance travelled for both the shoulder and the elbow when compared to visual feedback. This was an expected result since subjects could not visualize or know their current position as in the case of visual feedback and thus kept overshooting or undershooting their mark and had to compensate for their extra movement leading to an overall higher total displacement. As in the case of visual feedback, the users

TABLE III: The table below shows the mean results for the objective and subjective metrics. The audio and Multi-modal systems are individually compared to the Visual system to check for statistical significance. The values highlighted in green indicate a p value of < 0.05 which was obtained after carrying out a two tailed Student's T test.

Metric	Audio	Visual	Multi-modal
Task Completion Time (seconds)	40.97	22.48	22.28
Total Distance Travelled (degrees)	Shoulder = 134.53	Shoulder = 108.39	Shoulder = 119.62
	Elbow = 172.31	Elbow = 106.31	Elbow = 119.51
NASA - TLX Task load Index	42.16	41.49	28.99
Van der Laan	Usefulness = 1.27	Usefulness = 1.15	Usefulness = 1.54
	Satisfaction = 0.66	Satisfaction = 1.03	Satisfaction = 1.28

started off by moving their arms quickly to carry out large movements and as soon as they were close to desired position, their movements became controlled and careful. But with audio feedback, given the difficultly one has to process and implement angles mentally, users tended to move by larger angles even when they were very close to position leading to additional overall movement.

Multi-modal feedback on the other hand can be seen to have a slightly larger overall movement for either joint when compared to visual. One must keep in mind that in the Multi-modal feedback case, the audio feedback was activated only when the user was +- 5 degrees away from the desired position. However, users were seen to move by larger amounts because they found it a little difficult to approximate the angles. The additional adjustments one had to make attributed to the slightly larger angular displacements. One can however argue and say that with continued usage and practice, small angle approximation will get better for a user and this overall movement could further be reduced.

Another interesting thing that we can see from the graph (See Fig: 5) is that in the case of audio feedback, users can be seen to move their elbows more than their shoulders. This could be because of a number of reasons. Firstly, since the elbow is connected to the shoulder, while changing the relative position of the shoulder subjects unknowingly also changed the position of their elbow as well. Thus, when a user moves his/her shoulder by the said amount he/she also tends to move the elbow joint by a small amount without having received any command for elbow movement. This way subjects had to compensate for the extra elbow movement leading to a higher displacement of the elbow as compared to the shoulder. Thus, if subjects tried to keep the relative position of one joint constant while moving the other, this error could be avoided.

This problem cannot be seen in either of the other cases. This could be because in both visual and multi-modal the users rely on visual feedback more and thus can move both joints simultaneously in a coordinated manner while trying to reach the desired position and also they have constant real time feedback and this way they can approximate their errors better.

C. NASA-TLX

Audio and visual feedback seem to be almost equally demanding with no significant difference. The higher workload for audio can be attributed to frustration. Users reported extremely high ratings for this metric in particular as a constant audio feedback, while carrying out a task can be annoying, leading to a high frustration score across almost all participants. While in visual feedback, the effort metric was deemed the most challenging as users had to constantly look towards a screen while carrying out a task to check whether they were in the right position or not, leading to a larger effort to complete the task.

Multi-modal feedback on the other hand was seen to have a significantly lower overall workload as compared to visual. This was an expected

result since the multi-modal feedback specifically reduced the effort metric as users no longer had to look towards a screen to receive feedback but they could now concentrate on the task after having reached close enough to the desired position and carry out minute changes by only paying attention to the audio based feedback and had to shift focus to the screen only when larger position changes were needed. Further, the multi-modal system also significantly reduced the frustration scores as audio feedback was limited to only a small portion of the trial rather than be present at all times.

However, a large variance can be seen in the multi-modal feedback as compared to the other two because some users found the feedback to be slightly more demanding because there wasn't any alert mechanism to let them know when they needed to switch from one feedback modality to another. There were certain cases where subjects while receiving audio feedback overshot their mark and deviated from the $\pm$ 5 degree threshold and there was no way for them to know that they now had to switch to visual feedback to check and see by how far they had deviated. This led to them rating the Multi-modal feedback slightly higher as compared to the other participants and in turn leading to a larger variance in the overall results.

D. Van der Laan

All feedback modalities can be seen to be in the upper right quadrant indicating that they are acceptable to the users. However the multi-modal feedback was indicated to be more useful and satisfying as compared to the other two. The usefulness could be higher given that it helped redistribute the mental task workload by allowing users to concentrate more on the task rather than the feedback unlike visual feedback where users could either concentrate on the task or on the feedback. The satisfaction rating was also the highest because the system presented audio feedback only when users were close enough to the desired pose and not throughout the trial which was the case for the only audio based feedback which can be seen to have the lowest satisfaction rating as compared to the others.

Based on the above findings we can come to some important conclusions. Multi-modal feedback took almost the same amount of time for task completion as compared to visual and given that visual was considered to be the benchmark, multi-modal feedback was successful in objectively saying that it did not take users longer time to complete the task even though they had to shift focus from one modality to another. Secondly, when comparing the overall shoulder and elbow movement, multi-modal feedback had a slightly higher overall movement as compared to visual. But we also suggested that this performance could be greatly improved with a little bit of practice and familiarization. Finally, multi-modal feedback provided a significantly lower overall task load index as compared to visual and also provided a higher acceptance rating given that it has the advantages of visual feedback that is, real time error approximation for large scale movements and additionally it also incorporates the advantage audio has of redistributing mental task load by allowing users to concentrate more on the task while receiving the desired feedback. Keeping these results in mind and carrying out a few design modifications (alerts to let users know when to switch from one modality to another and a slightly more pleasant and natural voice command) we can say that multi-modal feedback comprising of audio and visual will greatly help improve task performance and user satisfaction for ergonomic applications.

V. EXPERIMENT 2

A second experiment was also conducted which would help motivate our choice for using audio feedback in joint space for our multi-modal feedback system. For this experiment we decided to compare user performance between audio feedback in joint space and audio feedback in endpoint space using the same metrics as mentioned above.

Through the experiment we did find that users took significantly lesser time to complete

the task in end point space and at the same time also had an overall lesser movement of either joint as compared to joint space feedback (Refer to Appendix D for details) however we also found one major limitation end point space had which made us choose joint space feedback over it.

End Point space feedback has a major drawback.

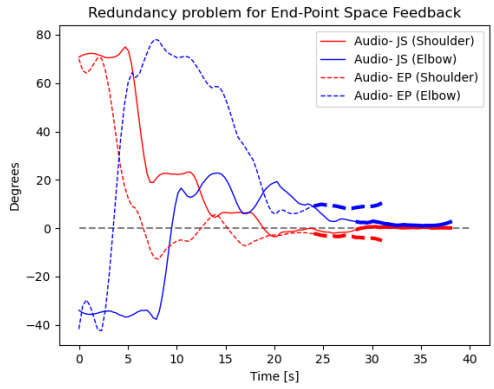


Fig. 7: The major drawback of using End-Point feedback can be very clearly seen from the figure above. The highlighted portion indicates the last 10 seconds of the trial where is user is asked to try and hold his position after having reached the desired pose. In Joint-Space feedback the user is successfully able to converge his joints to the desired pose whereas in End-Point feedback case, the user is far away from the desired joint position even though the endpoint (wrist) has successfully reached its desired position. This clearly shows its limitation for being used in an ergonomic guidance task in cases involving very high redundancy.

It cannot be used in cases where redundancy is high as it only focuses on the positioning of the endpoint and completely disregards the individual joint positions (See Fig:7). In other words, the users when receiving end point feedback can orient their joints in multiple ways to reach the same end point position. Thus, completely failing for an ergonomic guidance task which happens to be the main application of such a feedback modality. Joint Space feedback alone can be quite challenging for users but it does not suffer from redundancy problems. Thus, for cases of high redundancy one might prefer to receive audio commands in Joint Space making it the preferred audio feedback type for our multi-modal feedback system. End Point space feedback could be more useful and efficient for low redundancy movement cases such as movements limited to the lower body where one might assume the hip to the end point and the ankle joint to be the base.

VI. LIMITATIONS AND FUTURE WORK

A number of limitations were identified during our study. The pose estimation was carried out with a pre-trained model and thus had a lot of noise. This could be reduced by training the model further to suit our needs and in turn providing us with more accurate angular measurements. Some users reported the speech based feedback to be slightly unnatural and thus with continued usage this could lead to higher annoyance. The system is currently limited to carrying out pose estimation for only the right arm and thus further research needs to be done to expand its application to the whole body. Finally, the proposed feedback system is limited to quasi-static tasks such as drilling, polishing, grinding etc. where users have to hold their position for a few seconds to do the task. It would be of great interest to check the applicability of such feedback methods to real time dynamic tasks to further promote their need to prevent unnecessary musculoskeletal injuries.

VII. CONCLUSION

In this study we have successfully created a multi-modal feedback system for pose correction for an ergonomic task and tried to bridge the gap of a lack of direction based pose correction feedback mechanisms. The proposed feedback system was made up of a visual feedback which was used to carry out large range body movements and a speech based audio feedback system which was used for minor pose adjustments once the human was close enough to the desired pose. Through this feedback system we were successfully able to combine the advantages of visual and audio feedback namely, real time error approximation and redistribution of mental task load along with muti-dimensional information transmission which was done through the help of the speech based feedback. We had hypothesized that the proposed system would provide better performance as compared to the existing uni-modal feedback systems.

Our system showed equivalent performance

results when compared to the benchmark results of visual feedback when measured objectively through two metrics of task completion time and total distance moved. However, subjective results showed significantly lower overall task load index and better acceptance across users. Thus making a strong case for the need of a multi-modal feedback system for guiding humans from unergonomic to ergonomic poses.

Each modality was responsible for a particular perception and we believe that the proposed multi-modal system successfully combined these and we in turn created a more reliable and easy to use feedback method for position guidance.

REFERENCES

[1] Cao, Zhe and Hidalgo, Gines and Simon, Tomas and Wei, Shih-En and Sheikh, Yaser (2018),OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields, arXiv preprint arXiv:1812.08008

[2] Shrawan Kumar, Theories of musculoskeletal injury causation, Ergonomics, 2001, 44:1, 17-47, 10.1080/00140130120716

[3] Feuerstein, M.,In: Moon SD, Sauter SL, Definition, empirical support, and implications for prevention, evaluation, and rehabilitation of occupational upper extremity disorders, Beyond Biomechanics: Psychosocial Aspects of Musculoskeletal Disorders in Office Work. Bristol, PA: Taylor & Francis,177–206,1996

[4] National Research Council (US) and Institute of Medicine (US),Musculoskeletal Disorders and the Workplace: Low Back and Upper Extremities,Committee on Human Factors. Commission on Behavioral and Social Sciences and Education,Washington, DC: National Academy Press,2001

[5] Hagberg, M., Silverstein, B., Wells, R., Smith, M.J., Hendrick, H.W., Carayon, P. and Perusse, M., Work-Related Musculoskeletal Disorders (WMSDs):a Reference Book for Prevention, 1995,London: Taylor Francis,

[6] Kim, Wansoo and Lorenzini, Marta and Balatti, Pietro and Nguyen, Phuong DH and Pattacini, Ugo and Tikhanoff, Vadim and Peternel, Luka and Fantacci, Claudio and Natale, Lorenzo and Metta, Giorgio and others, Adaptable workstations for human-robot collaboration: A reconfigurable framework for improving worker ergonomics and productivity,,IEEE Robotics Automation Magazine,26,3,14–26,2019,IEEE

[7] Lorenzini, Marta and Kim, Wansoo and De Momi, Elena and Ajoudani, Arash, A Real-time Graphic Interface for the Monitoring of the Human Joint Overloadings with Application to Assistive Exoskeletons,International Symposium on Wearable Robotics,281–285,2018,Springer

[8] Portnoy, Sigal and Halaby, Orli and Dekel-Chen, Dotan and Dierick, Effect of an auditory feedback substitution, tactilo-kinesthetic, or visual feedback on kinematics of pouring water from kettle into cup,Applied ergonomics,51,44–49,2015,Elsevier

[9] Kim, Wansoo and Lorenzini, Marta and Kapıcıoğlu, Kağan and Ajoudani, Arash, Ergotac: A tactile feedback interface for improving human ergonomics in workplaces, ,IEEE Robotics and Automation Letters,3,4,4179–4186,2018,IEEE

[10] Sigrist, Roland and Rauter, Georg and Riener, Robert and Wolf, Peter, Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review,Psychonomic bulletin review,20,1,21–53,2013,Springer

[11] Elke Schneider and Xabier Irastorza, European Agency for Safety and Health at Work (EU-OSHA),OSH in figures: Work-related musculoskeletal disorders in the EU - Facts and Figures,Publications Office of the European Union,2010,978-92-9191-261-2, 10.2802/10952

[12] Carayon, P., Smith, M.J. and Haims, M.C., Work organization, job stress, and workrelated musculoskeletal disorders,Human Factors,1999,41:4,644-663,10.1518/001872099779656743

[13] Van der Laan, J.D., Heino, A., De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. Transportation Research - Part C: Emerging Technologies, 5, 1-10

[14] S. G. Hart and L. E. Stavenland. Development of Nasa-tlx (task load index): Results of empirical and theoretical research.Adv. Psychol, 52:139–, 12 1988

[15] Stephen Brewster, Nonspeech Auditory Output, The Human-Computer Interaction Handbook, Volume 1, Edition 2, 247-265

[16] Yinpeng Chen, He Huang, Weiwei Xu, Richard Isaac Wallis, Hari Sundaram, Thanassis Rikakis, Todd Ingalls, Loren Olson, Jiping He, The design of a real-time, multimodal biofeedback system for stroke patient rehabilitation,Proceedings of the 14th Annual ACM International Conference on Multimedia, MM 2006,763-772

Appendix A

This section explains the techniques used for estimating the pose and joint angular orientations in real time.

1.1 Pose Estimation

To carry out the pose estimation it was important to select a technique which would help us keep track of the kinematic positions of the joints while the human was doing the task. In order to successfully do so we decided to make use of an open source deep learning based pose estimation software, OpenPose [1].

OpenPose is an application that allows real time human pose estimation. It uses an RGB image as input for its multi-stage convolutional neural network. The neural network in turn produces a set of confidence maps and part affinity fields which are processed to give the 2D joint key points of the humans in the image based on the type of human body model used.

It currently has 3 types of body models Body-25, COCO and MPII. For our purposes, we have made use of the Body-25 model as it was the model with the best balance between accuracy and speed. The system outputs the coordinates of 25 joints of the human body based on the Body 25 model as shown in the figure below (See Fig: 1.1).

Since our experiments are limited to the right hand side of the body only, we were only concerned with the key points mentioned in the table below (See Table: 1.1).

Table 1.1: Extracted keypoints from the Body-25 human model

Key Point	**Joint Name**
2	Right Shoulder
3	Right Elbow
4	Right Wrist
9	Right Hip
10	Right Knee
11	Right Ankle

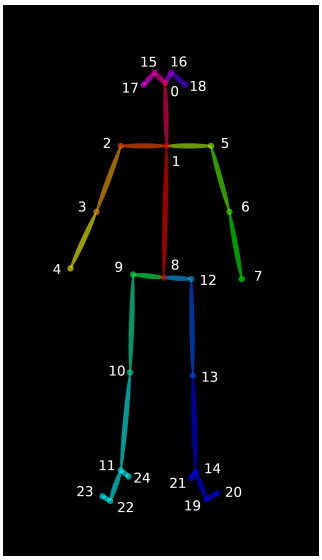


Figure 1.1: Body-25 model used for pose estimation.

1.1.1 Using a markerless motion capture technique for pose estimation

Traditional techniques of pose estimation or motion capture such as sensor based tracking (eg. using IMus) or marker based optical tracking (eg. using commercially available VICON or Optotrack) have not been used for our study. The reasons behind this have been explained ahead.

Use of a marker based optical tracking technique was rejected because such systems suffer from problems of (1) occlusion - where if a marker is covered by an object or other body part, the system is unable to detect the marker and this in turn leads to errors during data generation and (2) these systems also require a controlled laboratory environment involving fixed and calibrated camera setups, thus limiting their use in a dynamic environment.

Sensor based tracking techniques on the other hand involve attachment of multiple sensors on different anatomical key points which might hinder natural motion.

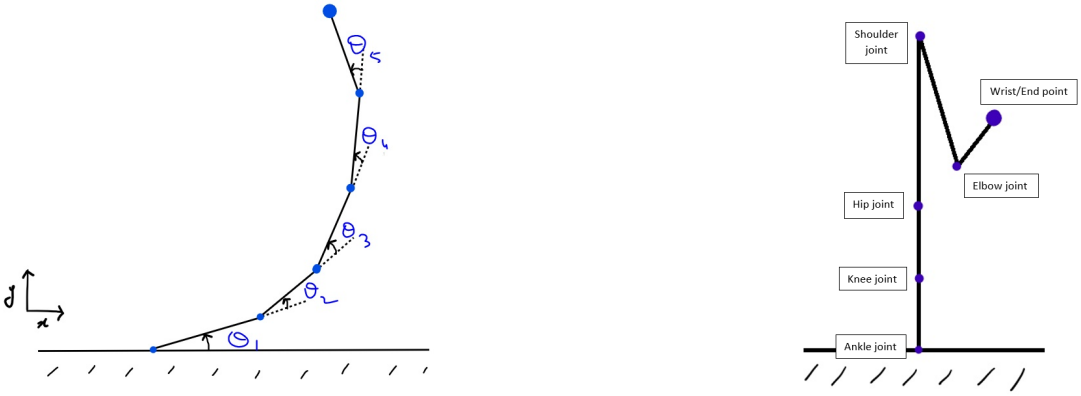
Based on the above mentioned limitations it was decided to make use of a markerless optical tracking technique to carry out pose estimation for our study as such systems do not suffer from problems of occlusion, controlled laboratory environments and also do not hinder natural motion. Further, studies such as Dutta[2],Clark et al.[3],Schmitz et al.[4] and Nakano et al. [5] also suggest that markerless optical tracking techniques are almost as accurate as the traditional techniques of motion capture.

Thus, based on the findings and conclusions a markerless optical tracking technique was made use of to carry out pose estimation for our study.

1.2 Body Model

After successfully extracting the real time key points coordinates it was important to find the real time angular positions of the joints since our feedback system was carrying out position guidance through angle based commands. In order to do so, a 5 degree of freedom human body model was created using the extracted joint key point coordinates. The body model is a stick figure representation of the human's right hand side viewed in the saggital plane.

The angular measurements of joints are found with respect to the previous joint as seen in the figure (See Fig: 1.2). We decided to treat each limb as a vector and found the angle made by one segment, relative to the previous segment.



(a) 5 DOF human body model where each joint angle is calculated in reference to the previous joint.

(b) Stick figure representation of the model visualized as a human with the labelled joints.

Figure 1.2: Body model

Appendix B

This section explains the methods used to create the different types of feedback systems used for our study.

2.1 Audio Feedback

Two types of Speech based audio feedbacks were created and used for this study. Both feedbacks were created using the Text-to-speech module in-built in the Python library.

2.1.1 Joint Space Feedback

In this type of feedback, commands were provided in a joint specific manner where the direction and magnitude of movement was provided. Since our study was limited to the the manipulation of the arms, only elbow and shoulder joint manipulations were carried out. At each iteration of the loop we checked the current and desired angular position of each of the joints and accordingly asked the users to move the joint by the differing amount. We assumed a threshold of $\pm$ 5 degrees for the joint angle positions. Thus the feedback always guided the user to positions within the defined threshold of the target position.

To calculate the angular orientation of a particular joint the coordinates of the current joint, and the joint preceding and proceeding it were found. The vector length between three segments was found using (See equations: 2.1, 2.2 and 2.3). Upon successfully obtaining the vector lengths between the three segments the angular orientation of the middle joint (current joint) was found with respect to the other two joints. And since we wanted the angular orientations of joints with respect to the previous joint, the final orientation was obtained by using (See equations: 2.4 and 2.5) .

$$p_1 = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{2.1}$$

$$p_2 = \sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2} \tag{2.2}$$

$$p_3 = \sqrt{(x_1 - x_3)^2 + (y_1 - y_3)^2} \tag{2.3}$$

Where p refers to the length of the vector and x and y are the joint keypoint coordinates.

$$z = \frac{(p_1^2 + p_2^2 - p_3^2)}{(2 * p_1 * p_2)} \tag{2.4}$$

$$\theta = \pi - cos^{-1}(z) \tag{2.5}$$

Where z is the angle made by the current joint with respect to the other two joints and θ denotes the final angular orientation of the current joint with respect to the previous joint.

2.1.2 End Point Space

When giving feedback in end point (Cartesian) space we assumed the wrist to be our end point and used the angular positions of the joints as inputs to a kinematic equation to receive the X and Y cartesian coordinates (See equations: 2.6 and 2.7) of our endpoint. The subject received speech commands which tried to guide them towards the desired coordinates of the end point and the unit of movement was in centimeters. The ankle joint was assumed to be the base and it was also assumed that the subject stands upright at all times, thus the ankle, knee and hip angles were assumed to be 90,0 and 0 degrees respectively (based on the human body model as explained before) whereas the shoulder and elbow angles were estimated in real time and continuously updated during the trial.

$$X = l_1.cos\theta_1 + l_2.cos(\theta_1+\theta_2) + l_3.cos(\theta_1+\theta_2+\theta_3) + l_4.cos(\theta_1+\theta_2+\theta_3+\theta_4) + l_5.cos(\theta_1+\theta_2+\theta_3+\theta_4+\theta_5) \tag{2.6}$$

$$Y = l_1.sin\theta_1 + l_2.sin(\theta_1+\theta_2) + l_3.sin(\theta_1+\theta_2+\theta_3) + l_4.sin(\theta_1+\theta_2+\theta_3+\theta_4) + l_5.sin(\theta_1+\theta_2+\theta_3+\theta_4+\theta_5) \tag{2.7}$$

Table 2.1: Symbols used for finding the kinematic coordinates of the end point and their meaning. Please also refer to Figure 1.2.

Symbol	**Meaning**
θ_1	Ankle Angle
θ_2	Knee Angle
θ_3	Hip Angle
θ_4	Shoulder Angle
θ_5	Elbow Angle
l_1	Distance between knee and ankle
l_2	Distance between knee and hip
l_3	Distance between hip and shoulder
l_4	Distance between shoulder and elbow
l_5	Distance between elbow and wrist

Since we had previously assumed a threshold of $\pm$ 5 degrees for each joint, we first calculated the end point coordinate positions for the upper and lower limits defined by these

thresholds. The feedback mechanism guided the subjects by instructing them to stay within the space bound by these coordinates.

2.1.3 Why use a speech based audio feedback?

Unlike the norm, where most feedback systems employing auditory feedback use non-speech auditory feedback techniques such as earcons, sonification etc., we decided to make use of a speech based feedback method. One of the main reasons for making such a choice is the ability of speech based feedback to provide a dynamic range of instructions. Given the multi-dimensional nature of the feedback commands needed to be provided for our study namely, joint name, direction of motion and magnitude of movement, using a speech based audio feedback technique seemed to be the preferred choice. Speech based feedback is also superior to other non-speech feedback techniques in cases where absolute values need to be transmitted [7] which happens to the magnitude of movement of the arm and its joints in our case. Speech based feedback is also free from the problem of habituation [7] which other non-speech audio feedbacks suffer from, where users get lost in the background because of the continuous presence of sound and are alerted only when the sound changes or stops.

2.2 Visual Feedback

To carry out visual feedback a stick figure representation of the human was made using OpenCV [6]. The desired pose was shown in white and the real time pose was shown in red. Since, the feedback was limited to the arm, we assumed that the subject would be standing upright throughout and thus the angles of the ankle, knee and hip were assumed to be such that the figure seemed upright.

The angular joint positions for the elbow and shoulder were used to plot lines using the equations shown below,

$$X_{(i+1)} = X_i - cos\theta_i \tag{2.8}$$

$$Y_{(i+1)} = Y_i - sin\theta_i \tag{2.9}$$

Where i refers to the joint previous (or below) to the current joint. Thus, the ankle angle would be used to find the coordinates of the knee and the knee angle would be used to find the coordinates of the hip and so on. Additionally, X and Y represent the pixel coordinates of the specified joints.

2.3 Multi-modal Feedback System

A multi-modal feedback system is one which has the ability to respond to a single input in multiple ways. For our study, the multi-modal system uses joint angle orientations as input and outputs direction based guidance through either visual or a speech based audio feedback in joint space.

Multi-modal systems have numerous advantages over other uni-modal systems [**?**] such as: increasing system robustness, increasing communication bandwidth between human and machine, increase the user's ability to correct errors, etc.

Our proposed multi-modal system processes each information channel of visual and audio separately and thus does not increase the cognitive load by providing both feedbacks in a concurrent manner. We believe that by processing these channels separately and using each channel to solve a different type of problem will in turn improve the quality of the solutions as compared to the solutions provided by uni-modal systems.

Our multi-modal system uses the visual channel to guide users to carry out large pose correction movements since providing real time error visualization is easiest to understand through a visual system and the audio system is activated only when users are close enough to the desired position and no longer need the help of a visual tool. This way, the cognitive overload that could be caused through a visual feedback system is avoided as users no longer need to look elsewhere for feedback and can concentrate on the task at hand.

Appendix C

This section explains the two subjective measures used to evaluate task performance for each of the experiments.

3.1 NASA Task Load Index

Subjects were asked to fill in a NASA-TLX questionnaire after undergoing 5 trials of the same condition in order to obtain the overload task load index. The questionnaire is used to evaluate the overall task load by taking a weighted average of six sub-scales namely, mental demand, physical demand, temporal demand, frustration, effort and performance. Out of these, the first three relate to the demands forced by the task on the subject and the latter three relate to the demands imposed due to the interaction between the task and subject[**?**].

The questionnaire is a two part form made up of weightings and ratings. Subjects are initially asked to fill in the weightings form (See Fig: 3.1a) which includes 15 pair wise comparisons of the six previously mentioned sub-scales. Subjects are asked to circle the metric which was more demanding for them between the two while doing the task. After receiving feedback for the 15 pairs, the frequency of the number of times a particular sub-scale is selected is noted which acts as a weighting factor for the overall task load evaluation. Each sub-scale can receive a weight between 0 to 5.

Subjects are then asked to fill in the ratings form (See Fig: 3.1b) where they are asked to approximate the workload contribution of each individual sub-scale while doing the task. The ratings are between 0 (not demanding at all) to 100 (extremely demanding).

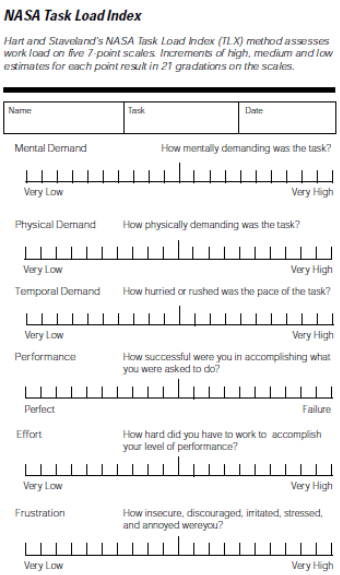
After obtaining the weighting and the rating for each sub-scale, the weight is multiplied to the rating to obtain the adjusted rating for each sub-scale. The adjusted ratings of the six sub-scales are summed together and divided by 15 to obtain the overall task load index.

3.2 Van Der Laan Questionnaire

Subjects were also asked to fill in the Van der Laan questionnaire to assess the acceptability of the feedback method through a two dimensional scaling system made up of Usefulness and Satisfaction [**?**]. The questionnaire involves a list of nine likert scaled questions (See Fig:

Effort Or Performance	Temporal Demand Or Frustration
Temporal Demand Or Effort	Physical Demand Or Frustration
Performance Or Frustration	Physical Demand Or Temporal Demand
Physical Demand Or Performance	Temporal Demand Or Mental Demand
Frustration Or Effort	Performance Or Mental Demand
Performance Or Temporal Demand	Mental Demand Or Effort
Mental Demand Or Physical Demand	Effort Or Physical Demand
Frustration Or Mental Demand	

(a) 15 pair wise comparison of the six sub-scales to obtain weights of individual sub-scales towards overall task load.



(b) Form to obtain raw ratings of each sub-scale while doing the task.

Figure 3.1: Paper-pencil version of the NASA-TLX

3.2) rated between -2 to +2. Usefulness is measured by summing the scores of questions 1,3,5,7 and 9 and dividing the total by 5 (end score between -2 to +2) whereas Satisfaction is measured by summing the scores of questions 2,4,6 and 8 and dividing the total by 4 (end score between -2 to +2). A graph can be plot between the usefulness and satisfaction scores and if the point lies in the top right quadrant, the system is deemed acceptable.

1 Useful	|__|__|__|__|__|__| Useless
2 Pleasant	|__|__|__|__|__|__| Unpleasent
3 Bad	|__|__|__|__|__|__| Good
4 Nice	|__|__|__|__|__|__| Annoying
5 Effective	|__|__|__|__|__|__| Superfluous
6 Irritating	|__|__|__|__|__|__| Likeable
7 Assisting	|__|__|__|__|__|__| Worthless
8 Undesirable	|__|__|__|__|__|__| Desirable
9 Raising Alertness	|__|__|__|__|__|__| Sleep-inducing

Figure 3.2: Nine likert scaled questions from the Van der Laan questionnaire.

3.3 Data Filtering

Before analysing the results of the experiments it was important to filter the data and remove any outliers which might have occured due to noise or sensor errors. We used a Savitsky-Golay filter having a window size of 11 and order 3 to make our data more readable and free off outliers (See Fig: 3.3).

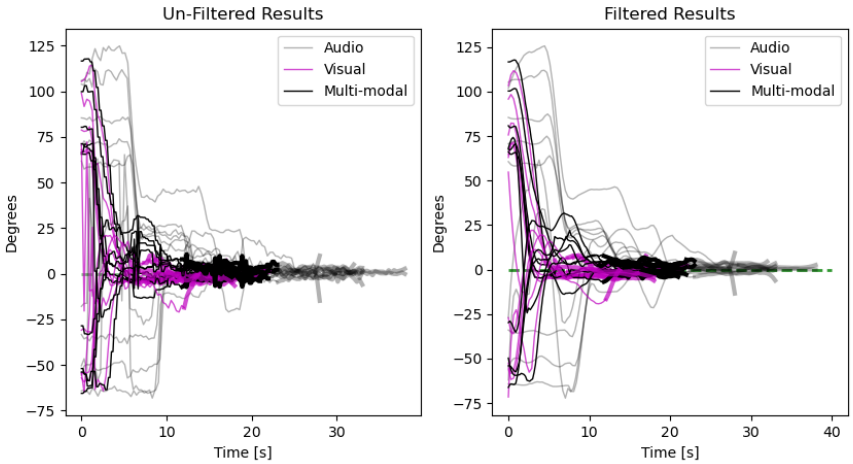


Figure 3.3: Plot showing the raw unfiltered v/s filtered data of all feedback modalities for one random subject. Each subject underwent 5 trials of the same condition.

Appendix D

This section explains the results of the experiment conducted to motivate our choice of using a speech based audio feedback in joint space.

4.1 Experiment 2

The goal of this experiment was to motivate our choice of using a speech based audio feedback in joint space for our proposed multi-modal feedback system.

4.1.1 Method

Subjects were asked to stand opposite to a camera system with their right hand side saggital plane facing the camera. Subjects were asked to stand in a neutral position (arm stretched out or arm straight over the head) and had to follow the instructions provided by the feedback mechanism to reach the desired pose. Upon successfully doing so a beeping noise was heard and subjects had to try and hold the end position for 10 seconds.

Subjects underwent 2-3 trial runs to familiarize themselves to the feedback method and were asked to complete the task as quickly as possible with minimum errors.

4.1.2 Results

Task Completion Time

Audio feedback in end point space (mean = 29.41 seconds, std deviation = 3.831 seconds) can be seen to have a significantly lower ($p<0.05$) task completion time as compared to joint space feedback (mean = 40.97 seconds, std deviation = 8.038 seconds)(See Fig: 4.1).

Total Distance Moved

No significant difference($p=0.6$) in overall joint movement can be seen for either joint in either conditions. For end point space feedback (shoulder - mean = 126.66 degrees, std deviation = 34.13 degrees; elbow - mean = 161.169 degrees, std deviation = 64.377 degrees) a slightly lower overall average for either joint can be seen as compared to joint space feedback

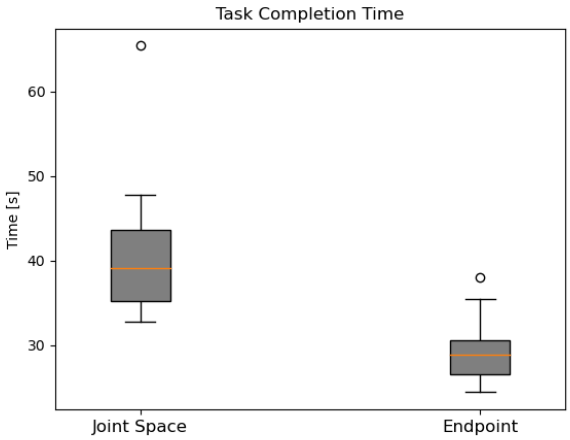


Figure 4.1: Box plot of the total task completion time for each audio based feedback condition can be seen above. Joint space feedback can be seen to have a higher task completion time as compared to audio feedback in end point space.

(shoulder - mean = 134.533 degrees, std deviation = 42.462 degrees; elbow - mean = 172.315 degrees, std deviation = 54.803 degrees)(See Fig: 4.2).

Subjective Measures

Audio feedback in joint space can be seen to have a mean overall task load index of 40.235 as compared to 43.996 for end point space feedback(p=0.53)

The results of the Van der Laan questionnaire show that the scores for both feedback mechanisms lie in the upper right quadrant, making them acceptable by users.

Additionally, we also asked subjects to rank their preferred feedback mechanism to further corroborate the results of the subjective questionnaires. Surprisingly, 10 out of 14 subjects preferred the audio commands in joint space (See Fig: 4.3).

Table 4.1: The table below shows the mean results for the objective and subjective metrics. The results of the audio feedback systems in joint space and end point space are compared through Student's T Test to check for statistical significance. The values highlighted in green indicate a p value of < 0.05.

Metric	Joint Space	EndPoint Space
Task Completion Time (seconds)	40.97	29.41
Total Distance Moved (degrees)	Shoulder = 134.53	Shoulder = 126.66
	Elbow = 172.31	Elbow = 161.17
NASA TLX	40.235	43.996
Van der Laan	Usefulness = 1.27	Usefulness = 1.13
	Satisfaction = 0.77	Satisfaction = 0.86

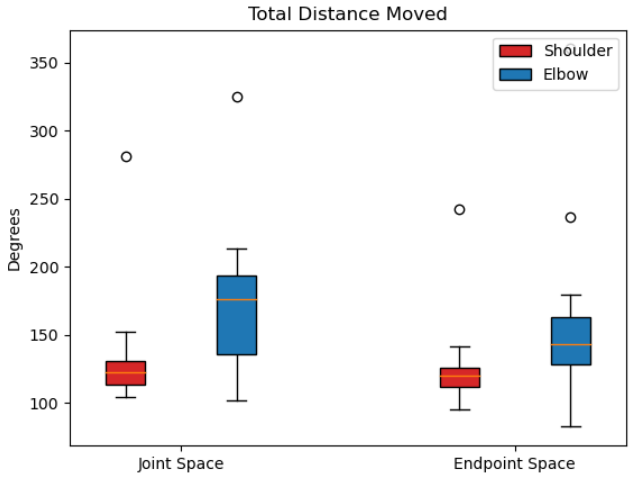


Figure 4.2: A box plot showing the total distance moved by the individual joints during the different audio feedback conditions.

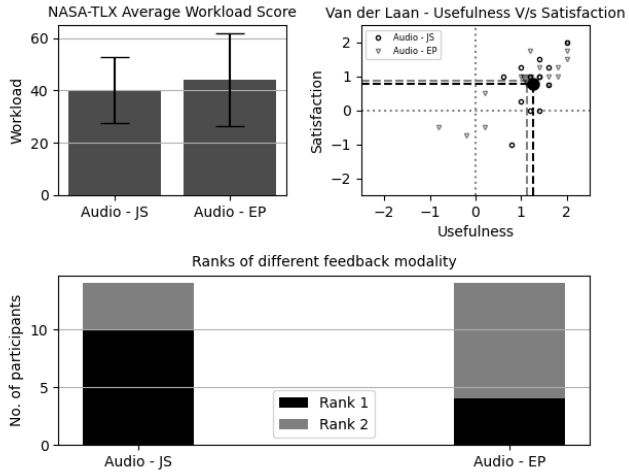


Figure 4.3: The figure above shows the results for the three subjective measures used to evaluate task performance. Both feedback types seem to be equally demanding but when asked to rank their preferences, more users seem to prefer audio feedback in joint space over endpoint space

4.1.3 Discussion

Task Completion Time

Audio feedback in endpoint space can be seen to have a significantly lower task completion time as compared to audio feedback in joint space. This was an expected result, as in joint space the user receives joint specific commands one after another and has to wait individually for every joint specific command to finish before carrying out the desired movement. Whereas in endpoint space, the user is asked to manipulate only his end point (wrist, in our case) and thus carries out coordinated movement of his other joints without having to wait for specific commands for them. Additionally, it is also easier to manipulate the arm as a whole rather than manipulate individual joints leading to smaller overshoots and thus lesser compensatory

corrective movements leading to an overall lesser time.

Total Distance Moved

No significant difference can be seen between the two conditions when comparing the total movement for each individual joint. Endpoint manipulation did show slightly lesser average movement when compared to joint space for each joint. This could be attributed to the fact that in end point space feedback asked the users to manipulate their entire arm and focused on positioning the endpoint making it easier for the user to carry out coordinated movements with lower over/under –shoots as they no longer had to carefully position and manipulate each joint individually.

One can also see a larger distance moved by the elbow as compared to the shoulder for either condition. This can be explained by the fact that when receiving joint space commands users tend to move their elbow along with their shoulders when receiving shoulder commands and in turn had to compensate for the extra movement leading to an overall larger displacement for the elbow. Whereas when subjects received endpoint space feedback, the elbow joint was moved along with the shoulder joint at all times and in particular to carry out minor adjustments or moving the forearm up or down. Thus leading to an overall higher displacement for the elbow.

Subjective Measures

A NASA-TLX was used to measure the overall task load index while performing the trials in the two types of audio based feedbacks. No statistically significant difference was found between the task load. Both tasks seemed to be equally taxing to the user. Based on the above results one might say that receiving only speech based audio feedback is probably not preferred by users and can be annoying or frustrating. Joint space feedback was seen to have higher scores for the mental demand and effort metrics as users found it difficult to carry out individual joint manipulations while having to approximate angles during movement while for endpoint feedback users reported higher frustration scores. Overall higher task load scores for endpoint feedback could be attributed to the fact that users found it less efficient as compared to joint space feedback as they now they were asked to manipulate both their joints in order to reach the desired endpoint position as compared to specific and efficient joint movements in the case of joint space feedback.

When comparing the user acceptability for both feedbacks, we can see that both present themselves in the upper right quadrant making each one acceptable to users. However, there seems to be no statistical difference between the scores for either scales. End Point feedback was seen to be slightly more satisfying to users probably because of having less specific and mentally taxing commands whereas Joint Space feedback seemed to be more useful due to its efficient joint specific movement commands.

4.1.4 Conclusions

From the above objective and subjective measures we can draw some conclusions. Joint Space feedback took a significantly longer time to complete the task and also had a slightly higher overall movement while being rated equally demanding on the NASA-TLX. Despite the above shortcomings, users still preferred to receive Joint Space commands simply because of the fact that they are much more efficient. However, one major limitation of end point feedback was also identified during the study, it cannot be used in cases where redundancy is high as it only focuses on the positioning of the endpoint and completely disregards the individual joint positions (See Fig: 4.4). Meaning, it fails to be successful in cases where multiple joint positions are possible to reach the same end point. Thus, completely failing for an ergonomic guidance task which happens to be the main application of such a feedback modality. Joint Space feedback alone can be quite challenging for users but it does not suffer from redundancy problems. Thus, for cases of high redundancy one might prefer to receive audio commands in Joint Space making it the preferred audio feedback type for our proposed Multi-modal feedback mechanism. End Point feedback could be more useful and efficient for low redundancy movement cases such as movements limited to the lower body.

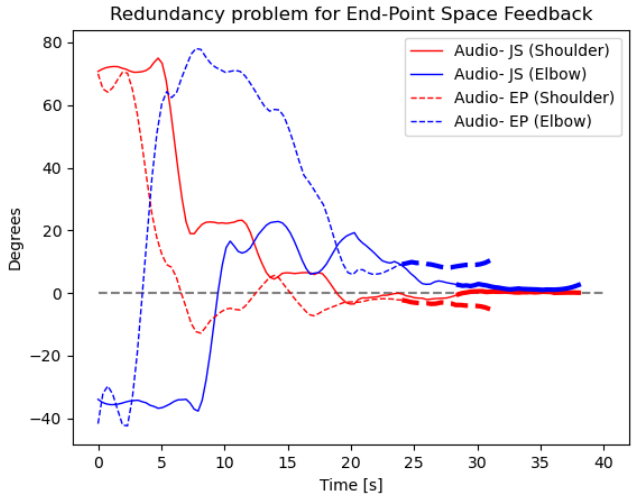


Figure 4.4: The major drawback of using End-Point feedback can be very clearly seen from the figure above. The highlighted portion indicates the last 10 seconds of the trial where is user is asked to try and hold his position after having reached the desired pose. In Joint-Space feedback the user is successfully able to converge his joints to the desired pose whereas in End-Point feedback case, the user is far away from the desired joint position even though the endpoint (wrist) has successfully reached its desired position. This clearly shows its limitation for being used in an ergonomic guidance task in cases involving very high redundancy.

Appendix E

5.1 Proposed Real Life Application

The main application of the proposed multi-modal system is to guide humans from unergonomic to ergonomic poses. A pose can be classified as unergonomic if during the pose certain biomechanical thresholds of joint torque overloads are crossed. Joint torques of individual joints can be found by making use of a human model similar to that used by [8] and further used by [9],[10],[11],[12], etc. The main objective of using such a model is to calculate the overloading joint torques when an external object was being handled by the human (eg. drilling machine, polishing machine, etc.). The model eliminates the use of any external sensors such as force plates to find ground reaction forces and calculates the overloading torques by finding the difference between the estimated Center of Pressure (CoP) between the human model with and without an external load.

The model uses angular orientation of the joints as an input to calculate the overloading torques. These orientations can be obtained by placing a stereo vision camera which will capture RGB images and use them as input for a deep learning based pose estimation software (eg. OpenPose [1]) which will provide the coordinates of the joint key points. The disparity of the stereo vision camera can further be used to calculate the 3D positions of the joint key points. Once joint key point positions are obtained one can use those to calculate the angular orientations.

Once the angular orientations are available, real time estimation of joint torque overloads can be obtained. Whenever certain thresholds are crossed, users can be alerted and guided to optimized positions where their torques are minimized using the proposed multi-modal system. The current system is for now limited to quasi-static tasks such as drilling, polishing, grinding, etc. where one needs to hold a pose for 5-10 seconds.

Provided certain design modifications are made to the feedback system to include real time position optimization and other changes as mentioned in the paper previously, we believe that the system will be extremely beneficial in preventing work related musculoskeletal injuries.

References

[1] Cao, Zhe and Hidalgo, Gines and Simon, Tomas and Wei, Shih-En and Sheikh, Yaser (2018),OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields, arXiv preprint arXiv:1812.08008

[2] Dutta, Tilak, Evaluation of the Kinect^{TM} sensor for 3-D kinematic measurement in the workplace, Applied ergonomics, 43, 4,645–649, 2012, Elsevier

[3] Clark, Ross A and Pua, Yong-Hao and Fortin, Karine and Ritchie, Callan and Webster, Kate E and Denehy, Linda and Bryant, Adam L, Validity of the Microsoft Kinect for assessment of postural control, Gait & posture,36,3,372–377,2012,Elsevier

[4] Schmitz, Anne and Ye, Mao and Boggess, Grant and Shapiro, Robert and Yang, Ruigang and Noehren, Brian,The measurement of in vivo joint angles during a squat using a single camera markerless motion capture system as compared to a marker based system, Gait & posture, 41,2,694–698, 2015,Elsevier

[5] Nakano, Nobuyasu and Sakura, Tetsuro and Ueda, Kazuhiro and Omura, Leon and Kimura, Arata and Iino, Yoichi and Fukashiro, Senshi and Yoshioka, Shinsuke, Evaluation of 3D markerless motion capture accuracy using OpenPose with multiple video cameras,bioRxiv,842492, 2019, Cold Spring Harbor Laboratory

[6] Bradski, G. (2000), The OpenCV Library, Dr. Dobb39, Journal of Software Tools

[7] Stephen Brewster, Nonspeech Auditory Output, The Human-Computer Interaction Handbook, Volume 1, Edition 2, 247-265

[8] Peternel, L., Kim, W., Babič, J., Ajoudani, A. (2017, November). Towards ergonomic control of human-robot co-manipulation and handover. In 2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids) (pp. 55-60). IEEE.

[9] Kim, W., Lorenzini, M., Kapıcıoğlu, K., Ajoudani, A. (2018). Ergotac: A tactile feedback interface for improving human ergonomics in workplaces. IEEE Robotics and Automation Letters, 3(4), 4179-4186.

[10] Kim, W., Lee, J., Peternel, L., Tsagarakis, N., Ajoudani, A. (2017). Anticipatory robot assistance for the prevention of human static joint overloading in human–robot collaboration. IEEE robotics and automation letters, 3(1), 68-75.

[11] Kim, W., Lorenzini, M., Balatti, P., Nguyen, P. D., Pattacini, U., Tikhanoff, V., ... Ajoudani, A. (2019). Adaptable workstations for human-robot collaboration: A reconfigurable framework for improving worker ergonomics and productivity. IEEE Robotics Automation Magazine, 26(3), 14-26.

[12] Lorenzini, M., Kim, W., De Momi, E., Ajoudani, A. (2018). A synergistic approach to the real-time estimation of the feet ground reaction forces and centers of pressure in humans with application to human–robot collaboration. IEEE Robotics and Automation Letters, 3(4), 3654-3661.