

Evaluation of Auditory Feedback on Task Performance in Virtual Assembly Environment

Ying Zhang

Department of Computing Science
University of Glasgow
17 Lilybank Gardens
Glasgow G12 8QQ
UK
Tel: +44 (0)141 3302333
Fax: +44 (0)141 3304913
Email: ying@dcs.gla.ac.uk

Reza Sotudeh

Department of ECEE
University of Hertfordshire
College Lane, Hatfield
Herts AL10 9AB
UK
Tel: +44 (0)1707 284154
Fax: +44 (0)1707 284199
Email: r.sotudeh@herts.ac.uk

Abstract

This paper presents our approach for the integration of auditory feedback into virtual assembly environment (VAE), and the investigation of the effect of auditory and visual feedback on the assembly task performance in virtual environments (VE). This VE experimental system platform brought together complex technologies such as constraint-based assembly simulation, optical motion tracking technology, and real-time 3D sound generation technology around a Virtual Reality (VR) workbench and a common software platform. Several experiments have been conducted to explore and evaluate the effectiveness of neutral, visual, auditory and integrated feedback mechanisms on task performance in the context of assembly simulation in VEs. Peg-in-a-hole assembly task has been used as the task case to perform the experiments, using sixteen subjects. Both objective performance data (task completion time) and subjective opinions (questionnaires) on the utilisation of auditory and visual feedback in VAE were gained from the experiments. The results showed that the addition of auditory feedback did introduce an improvement in the virtual assembly task performance. They also indicated that the statistically significant effect of the combination of auditory and visual feedback on the assembly task performance was better than individual feedback mechanism used alone. Most of the users preferred the combined feedbacks to the individual ones. The subjects' comments demonstrated that non-realistic feedback would defer performance and increase the level of frustration.

1. Introduction

In the manufacturing industry arena, Virtual Environment (VE) technology provides a useful method to interactively evaluate assembly-related engineering decisions, and to factor the human elements and considerations into finished products very early in the development cycle [1]. This could potentially lead to lower cost, higher product quality, shorter time-to-market, and improve competitiveness of the innovative products. Assembly is an interactive process involving the operator (user) and the handled objects, and hence simulation environments must be able to react according to the user's actions in real time. Furthermore, the action of the user and the reaction of the environments must be presented in an intuitively comprehensible way. Therefore, it is of great importance to investigate the factors related to information presentation modalities and effective mechanisms, which affect the human performance in performing assembly task in VEs. The multi-modal information presentation, integrated into the virtual environment, has potential for stimulating different senses, increasing the user's impression of immersion, and the amount of information accepted and processed by the user's perception system. Consequently, the increase of useful feedback information may enhance the operator's efficiency and performance. However, despite of recent efforts in assembly simulation [2,3] and 3D sound performance modelling in VE [4,5,6,7,8], very limited research has been performed to investigate and evaluate the effect of multi-modal feedback mechanisms, especially 3D auditory and visual feedback, on virtual assembly task-performance within virtual environment [9]. This paper presents the

overall system architecture implemented for creating a multi-modal virtual assembly environment, the approaches adopted to evaluate the factors affecting the user performance in performing the assembly tasks, the evaluation experiment and the relevant experimental results. It addresses (a) whether the integration of 3D auditory and visual feedback mechanisms can improve the assembly task performance more than the individual one within virtual environments; (b) which type of the feedback is the best among neutral, visual, 3D auditory and integrated feedback mechanisms.

2. Experimental Platform of the Assembly Task Performance

The hardware configuration and software architecture of the experimental system platform for multi-modal virtual assembly task performance, the schemes of auditory feedback generation, and unification of visual and auditory feedback presentation are addressed in this section.

2.1. Hardware configuration of the platform

The hardware configuration of the experimental system platform for virtual assembly task performance is comprised of three major parts: visualisation subsystem, auralisation subsystem, and the real-time optical motion tracking system (see Fig. 1). The core of the visualisation subsystem was a Trimension's V-Desk 6 fully integrated immersive L-shaped responsive workbench driven by Silicon Graphics Incorporated (SGI) desk-side Onyx2 supercomputer with four 250MHz IP27 processors and an InfiniteReality-2E Graphics board. The Trimension's V-Desk 6 was integrated with StereoGraphics' Crystal Eyes3 liquid crystal shutter glasses and the infrared emitter that connected to the Onyx2 workstation. These were used to generate stereoscopic images of the virtual world; one from the viewer's left eye perspective, and another one from the right eye. When the user used a Crystal Eyes liquid crystal shutter glasses to view the virtual world, these images were presented to the corresponding eye, and provided the user depth cues that made the immersive experience realistic.

The auralisation subsystem is based on a sound server (Huron PCI audio workstation), which was a specialised digital signal processing system. It employed a set of TCP/IP protocol-based procedures in terms of Spatial Network Audio Protocol (SNAP) to allow the virtual environment host to transmit the attributes of the assembly scene, positional information of the user and the sound-triggering event to the sound

server through a local area network (LAN). The VE host sends packets specifying the auditory-related attributes of the scene, and events such as collisions and motions between the manipulated objects, including the position of the event, and the user and environmental attributes, derived from the geometry of the assembly environment. From these packets, the auralisation subsystem generates a set of auralisation filters and sends them to the DSP boards. Based on an event-driven scheme for the presentation of objects' interactions, the DSP board samples and processes sound materials (data streams) with specified filters. Processed sound materials are then sent back to a set of headphones or an array of loudspeakers within the VE area in analogue form through coaxial cables. The auditory feedback in these experiments was presented to the user using a pair of the Sennheiser HD600 headphone.

The optical motion tracking system (Vicon's 612 workstation) provides dynamic, real-time measurement of position (X, Y and Z) and orientation (Azimuth, Elevation, and Roll) of the tracked targets such as head, hands, and manipulation tools, using passive-reflective markers and high speed, high resolution cameras. It was connected to the VE host using the TCP/IP protocol over a local area Gigabit Ethernet. A wand was used to support interactive object selection and virtual assembly operations. A virtual 3D pointer with ray-casting and a virtual hand were utilised as the interaction metaphor for the assembly operation.

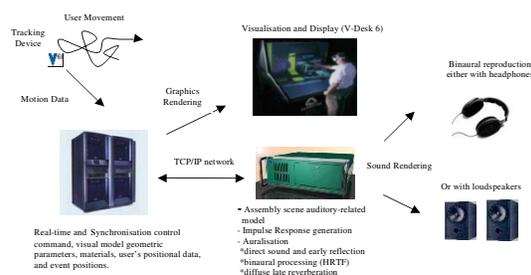


Fig. 1 Infrastructure of the System Platform

2.2. Software architecture of the platform

The software environment is a multi-threaded system that runs on SGI IRIX platforms. It consists of a *User-Interface/Configuration Manager*, *World-Manager*, *Input-Manager*, *Viewer-Manager*, *Sound-Manager*, *Assembly-Simulator*, *CAD Translator* and *CAD Database* (see Fig. 2). The *User-Interface/Configuration Manager* tracks all master processes to allow run time configuration of different modules.

The *World-Manager* is responsible for the administration of the overall system. It coordinates the visualisation, user inputs, databases, assembly simulation, and sound management. The *World-Manager* fetches the user inputs for manipulation, produces constrained motion using the *Assembly-Simulator*, and passes the corresponding data (e.g. the position and orientation information of the objects and the user) to the *Viewer-Manager* and *Sound-Manager* for auditory and visual feedback generation. The new data is used to update the scene graph and control the sound server via the *Sound-Manager*. The *World-Manager* also has the responsibility to synchronise various threads such as rendering and collision detection. Extensions to the OpenGL Optimizer have been made to view the scene using different display technologies (e.g. L-shaped Workbench, CAVE and Reality Room). The *Viewer-Manager* renders the scene to the selected display in the appropriate mode. Rendering is performed using parallel threads to provide real time response.

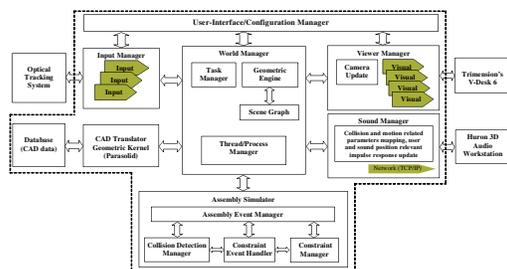


Fig. 2 Software Architecture

The *Input-Manager* manages user-object interactions and establishes the data flow between the user inputs and the objects held by the *World-Manager*. It supports devices such as pinch gloves, wands and Vicon's optical motion tracking system. These inputs describe the user actions/commands in the VE. Each device has its own thread to process data from it. These threads run in parallel with the rendering threads to achieve low latency. Once the assembly components are loaded into the scene graph via the *CAD-Translator*, the *Input-Manager* allows the user to select and manipulate objects in the environment. The *Sound-Manager* gets the location data for the user (listener/viewer), the positions of the collisions and motions (sound sources), and the parameters relating to sound signal modulation from the *World-Manager* and *Assembly-Simulator*, and then uses the Huron API to manage the audio workstation via local network using the TCP/IP protocol.

The *Assembly-Simulator* carries out the detection of collisions between the manipulated object and the surrounding objects, and supports interactive constraint-based assembly operations. During object manipulation, the *Assembly-Simulator* samples the position of the moving object to identify new constraints between the manipulated object and the surrounding objects. Once new constraints are recognized, new allowable motions are derived by the *Assembly-Simulator* to simulate realistic motion of assembly components. Parameters such as the accurate positions of the assembly components are sent back to the *World-Manager*, which defines their precise positions in the scene. When a constraint is recognised, the matching surfaces are highlighted to provide visual feedback, and/or 3D auditory feedback is generated through the *Sound-Manager* and the sound server.

For the virtual assembly scene management and rendering, the system utilises standard polygonal rendering techniques to visualise the assembly components and assembly scene. It has been implemented by the integration of the Parasolid geometric kernel and the OpenGL Optimizer graphical toolkit. The initial loading of the CAD models to the virtual world is performed via the Parasolid geometric kernel. Once the models are loaded, extracting polygons for each surfaces of each assembly component creates an Optimizer scene graph. In the virtual assembly scene representation, this system platform assumes that the user is always within the space termed *user space* and his/her body position is tracked with respect to the centre of the *user space*. The user may move the *user space* within the virtual world space using a navigational device, a wand in this system platform. The transformation node at the top of the *user space* allows the user to move the *user space* anywhere in the virtual world. The current *user space* structure maintains the head and the hand(s) positions within the *user space*. The Vicon's optical tracker values are assigned directly to the hand(s) and head transformation nodes to maintain their correct position within the virtual world. The virtual assembly scene graph has two kinds of spaces, the *background world space* and *assembly components space*, the *background world space* is maintained to provide a visually realistic assembly scene. The *background world space* is rendered but not available for interaction. The scene auditory-related models are associated with this space as well. The *assembly component space* maintains the assembly components, which are to be evaluated for assembly operation.

2.3. Auditory feedback rendering

Since the interaction of the users with the VE and the assembly simulation require sufficient real time behaviors, limited computation time is available for 3D sound generation. Due to the limitation of the available processing time, detailed auralisation has not been implemented in this work. In order to implement real time 3D sound within the limited computation power, some tradeoffs have to be adopted. In this research, binaural impulse responses are used to simulate the auditory-related attributes of the assembly scene and headphones are used to display the auditory feedback. The direct sound plus first order reflections are calculated in real time and mixed to form the B-format soundfield signals. The gain values used in this mixing are computed from the direction of arrival of each sound image at the user position and the parameters from the events such as the collision strength.

For the impulse response generation of the virtual assembly scene, the simplified image sources method has been utilised to calculate the room impulse response, and B-format has been selected to represent the soundfield. A box is used to approximate the volume of the geometry of the virtual assembly scene. The direct sound and the first-order reflections from six surfaces are calculated at run time. Each sound arrival (direct or reflected) is characterised by its time of arrival based on the distance travelled by the echo-path, its direction of arrival and level of attenuation due to either distance of sound propagation or material properties of the reflective surface. From the second order reflections to the reverberant tail of the impulse responses are pre-computed depends on the environmental parameters such as geometry of the scene, materials of the scene boundary, locations/orientations of the sound sources/users etc.

For soundfield representation and headphone display, the B-format is used as an intermediate format and then the B-format signals are decoded for headphone display since it is a convenient method for creating and manipulating 3D sound in auralisation systems. The B-format is referred to as *Ambisonics* and is essentially a four-channel audio format that can be recorded using a set of four coincident microphones arranged to provide one omni-directional channel (the W channel) and three figure-8 channels (X, Y, and Z channels). This set of X, Y, Z and W signals represent a first-order approximation to the sound field at a point within the assembly scene. The first step in the headphone display process, a DSP function is built to filter the four B-format components and produce two outputs, in such a way that a static binaural

presentation can be made of the B-format soundfield. The next step is to add a mixer that can rotate the X, Y, Z components of the soundfield prior to the binaural filters so that, in conjunction with the optical head-tracking device (Vicon's optical motion tracking system), the sound field remain static when the user rotates his/her head. The head tracking is achieved by rotating the X, Y, Z signals using a 3 x 3 matrix. The HRTF data are loaded from sound server disk to the DSP memory at run time.

2.4. Unification of visual and auditory presentation

The whole system platform consists of two major parallel data streams: the auditory stream and visual stream. The two streams share some common control and synchronisation mechanisms, and information sources. The outputs of the streams are what the user hears and sees within the semi-immersive virtual assembly environment. The visual aspects focus on the geometry definition, motion description, and the physical properties of the assembly components and visual feedback such as modification of colour, hue and saturation. The auditory part focuses on the mono and 3D auditory feedback generation process from sound activation, sound synthesis to sound propagation in the virtual assembly scene to auralisation in the ears of the user. The virtual world software extracts the spatial coordinates of the user and sound source positions that are transmitted via TCP/IP packets to a sound server (auditory stream) that runs a separate world model with the required auditory-related properties. The sound server then spatialises the sound materials according to the geometry information received and introduces the relevant scene attributes related to the auditory. The system components and the overall information flow are shown in Fig. 3. The upper half of the figure shows the auditory stream while the lower half shows the visual stream. Visual models are created using CAD tools, transformed and imported into this system with OpenGL Optimizer software. Auditory-related models are generated using CATT-Acoustic software and then loaded into Huron audio workstation with the relevant API.

3. Task-Performance evaluation

This section presents the experiments of assembly task performance evaluation including experiment hypotheses, objective evaluation, subjective evaluation, the relevant results and statistical analysis.

For the scenario of the VAE system platform, the user is presented with an assembly scene where the virtual components are initially located. The user then can perform assembly tasks, take decisions, make design changes, and perform a host of other engineering tasks in the VE. This research used two kinds of task cases to objectively (quantitatively) and subjectively (qualitatively) evaluate the effect of auditory and visual feedback on the assembly task performance respectively, it hypothesised that the performance would differ significantly between different feedback conditions. The performance was measured on the basis of objective and subjective means: a) objective means was the time taken to complete the assembly task, and b) subjective means was questionnaires and post-questionnaires for subjective ratings and preferences. The independent variables in these experiments are the feedback conditions of the multi-modal virtual assembly environment system as described in Table 1, namely, condition A, condition B, condition C and condition D. The dependent variable is the Task Completion Time (TCT) of each experiment condition, and subjective ratings and preferences.

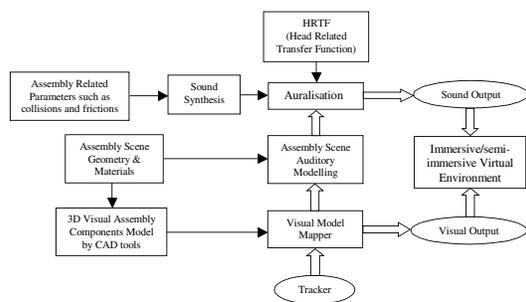


Fig. 3 System Information Flowgraph

Condition A	Colour	Sound
Condition A	× (Absent)	× (Absent)
Condition B	√ (Present)	× (Absent)
Condition C	× (Absent)	√ (Present)
Condition D	√ (Present)	√ (Present)

Table 1 Four Experimental Conditions

3.1. Experiments hypotheses

There were several hypotheses related to the experiments:

- **The use of visual feedback can lead to better performance and subjective satisfaction than neutral condition.** Task performance was measured by task completion time and subjective satisfaction. Task completion time was expected to decrease by providing essential collision, interaction and constraint cues by visual feedback for the assembly task. The subjective preference to and satisfaction with the interface with visual feedback was higher than with no additional feedback. It was expected that this would be shown by the visual feedback condition having faster task completion time and having statistically significant higher scores on the rating scales by the questionnaires as compared to the neutral condition.
- **The use of 3D auditory feedback can lead to better performance and subjective satisfaction than neutral condition.** Better performance would be shown by faster task completion time for the auditory feedback condition than the neutral condition. Auditory cues would provide more information for producing a realistic and productive application than no additional sensory cues, and the user could be better immersed with this information. The subjective preference to and satisfaction with the interface with auditory feedback was higher than with no additional feedback. It was expected that this would be shown by the auditory feedback condition having statistically significant higher scores on the rating scales related to ease of use as compared to the neutral condition.
- **The use of integrated feedback (visual plus auditory feedback) can lead to better performance and subjective satisfaction than any individual feedback mechanism used alone.** It was anticipated that this would be shown by faster task completion time and statistically significant differences between the related rating scale results for the integrated feedback as compared to the conditions with just auditory or visual cues.

3.2. Objective evaluation

For the objective (quantitative) evaluation, a peg-in-a-hole assembly task (for the scenario see Fig. 4, for the implementation see Fig. 5), which is relatively simple but relatively accurate for Task Completion Time (TCT) measurement, is used to explore and evaluate the effectiveness of neutral, visual, auditory

and integrated feedback mechanisms on the assembly task performance. The peg-in-a-hole assembly task has several phases (a) Placement of the peg to the upper surface of the plate (see Fig. 4a); (b) Collision between the bottom surface of the peg and the upper surface of the plate (see Fig. 4b); (c) Constraint recognition (see Fig. 4b); (d) Constrained motion on the plate (see Fig. 4c); (e) Alignment constraint between the peg cylinder and the hole cylinder (see Fig. 4d); (f) Constrained motion between two cylinders (see Fig. 4e); (g) Collision between the bottom surface of the peg ear and the upper surface of the plate (see Fig. 4f); and (h) Constraint recognition (see Fig. 4f). The different realistic 3D localised sounds or/and colour intensity of the colliding polygons are presented as the action cues for each of the aforementioned phases.

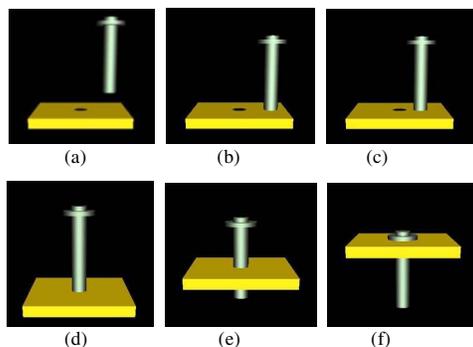


Fig. 4 Virtual Assembly Scenario of Peg-in-a-hole Task

This experiment is a 2×2 (two-factor) within-subjects design with the auditory feedback (present or absent) along with visual feedback (present or absent) being the within factors. For the four within factors (auditory \times visual), the presentation order was counterbalanced across subjects and conditions, and determined by employing 4×4 Latin Square, providing 16 different orders of feedback presentation. Each subject was randomly assigned to one of the orders. Although under each condition each subject went through 4 trials, only data of the assembly task completion time from the third and fourth trials were recorded to calculate the average task completion time under each condition and quantitatively analysed. The mono and 3D auditory were randomly across the subjects instead of the above auditory factor.

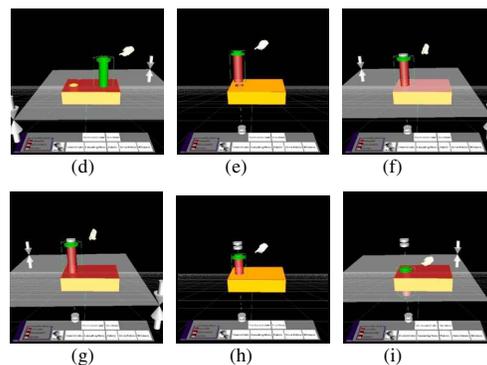
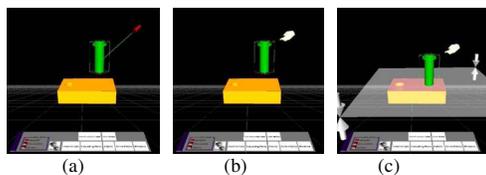


Fig. 5 Virtual Assembly Process and Feedback of Peg-in-a-hole Task

Sixteen subjects from the students and staff of the Centre for Virtual Environments in the University of Salford were invited to attend these experiments. All of them have normal or corrected normal visual acuity, normal colour vision and normal hearing. The task completion times, which represented the time span between the start and the end of the peg-in-a-hole task, were recorded by the software. The software timer was set to start, when the subject grabbed the peg to begin the assembly task progress. The software timer was set to stop, when the subject completed the assembly progress and released the peg. The system clock drove the timer. These data are summarised in Table 2 and illustrated in Fig. 6.

Both *two-way repeated measures ANOVA* and post-hoc pair-wise *t-test* comparisons were conducted on the task completion times in order to find the effects of four feedback conditions on the task performance respectively. *Two-way repeated measures ANOVA* on task completion times generated statistically significant results for auditory feedback ($F_{(1,15)} = 55.42 > F_{critical-value} = 4.54$ ($p < 0.05$)) and visual feedback ($F_{(1,15)} = 52.98 > F_{critical-value} = 4.54$ ($p < 0.05$)) respectively.

Conditions	Number of subjects	Mean (seconds)	SD
No feedback (neutral)	16	4.76	0.83
Visual	16	4.21	0.77
Auditory	16	4.19	0.81
Visual and Auditory	16	3.73	0.77

Table 2 Mean Completion Times

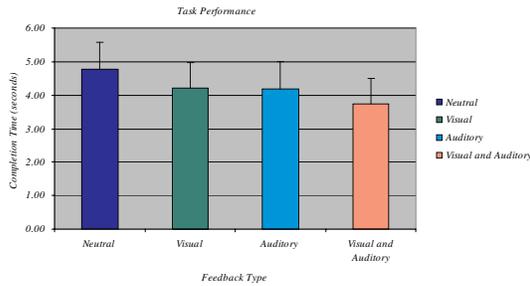


Fig. 6 Task Performance (mean completion times)

Pair-wise *t*-test comparisons of task completion times were conducted between neutral feedback condition, visual feedback condition, auditory feedback condition, and visual plus auditory feedback condition respectively. There were statistically significant differences between neutral feedback condition and visual feedback condition ($\mu_{\text{none-feedback (total)}} = 76.19\text{s}$; $\mu_{\text{visual-feedback (total)}} = 67.32\text{s}$; $t_{(15)} = 6.67 > t_{\text{critical-value}} = 2.13$ ($p < 0.05$)), between neutral feedback condition and auditory feedback condition ($\mu_{\text{none-feedback (total)}} = 76.19\text{s}$; $\mu_{\text{auditory-feedback (total)}} = 67.10\text{s}$; $t_{(15)} = 7.22 > t_{\text{critical-value}} = 2.13$ ($p < 0.05$)), between neutral feedback condition and visual plus auditory feedback condition ($\mu_{\text{none-feedback (total)}} = 76.19\text{s}$; $\mu_{\text{visual+auditory-feedback (total)}} = 59.63\text{s}$; $t_{(15)} = 8.23 > t_{\text{critical-value}} = 2.13$ ($p < 0.05$)), between visual feedback condition and visual plus auditory feedback condition ($\mu_{\text{visual-feedback (total)}} = 67.32\text{s}$; $\mu_{\text{visual+auditory-feedback (total)}} = 59.63\text{s}$; $t_{(15)} = 6.23 > t_{\text{critical-value}} = 2.13$ ($p < 0.05$)), between auditory feedback condition and visual plus auditory feedback condition ($\mu_{\text{auditory-feedback (total)}} = 67.10\text{s}$; $\mu_{\text{visual+auditory-feedback (total)}} = 59.63\text{s}$; $t_{(15)} = 6.51 > t_{\text{critical-value}} = 2.13$ ($p < 0.05$)) (see Table 5.5). There was no statistically significant difference between visual feedback condition and auditory feedback condition ($\mu_{\text{visual-feedback (total)}} = 67.32\text{s}$; $\mu_{\text{auditory-feedback (total)}} = 67.10\text{s}$; $t_{(15)} = 0.22 < t_{\text{critical-value}} = 2.13$ ($p < 0.05$)) (see Table 3). There was no obvious difference in the task completion time between the conditions of 3D localised and mono auditory feedback. In addition to these results, informal observation of frustration was much more frequent when subjects had no task feedback.

	None-feedback condition	Visual feedback condition	Auditory feedback condition	Visual + Auditory feedback condition
None-feedback condition		6.67	7.22	8.23
Visual feedback condition	6.67		0.22	6.23
Auditory feedback condition	7.22	0.22		6.51
Visual + Auditory feedback condition	8.23	6.23	6.51	
Critical Value	2.13 ($p < 0.05$, $df = 15$, two-tailed test) <i>t</i> -Value exceeds this value means the difference is highly significant			

Table 3 t-value for Pair-wise t-test Comparisons

3.3. Subjective evaluation

For the subjective evaluation of neutral, visual, auditory and integrated feedback mechanisms on the assembly task performance, a Sener electronic box and its brackets maintenance assembly task (see Fig. 7) was used. The difference with the objective evaluation is that this evaluation used the questionnaires and post-questionnaires to perform the subjective measurements including 10-point rating scales to evaluate the overall satisfaction, the realism, perceived task difficulty and performance, ease learning, perceived system speed and overall reaction to the received feedback. Additionally, after the subjects completed the tasks under all conditions they were required to complete a set of 7-point rating scales and open-ended questions comparing the different feedback cues. The 7-point rating scales asked the subjects to compare how well the different feedback cues helped them to complete the task, how they foresaw these cues helpful in a real design application, and which kind of feedback cues they preferred. Preferences were determined by asking subjects to rank the four conditions in the order of his/her preference when all trials were completed. Finally, subjects were asked to provide general opinions and comments about their experiences. The answers of the subjects were recorded and analysed. Table 4 shows the number of subjects who placed the

different conditions as first, second, third and fourth in their rankings. Fig.8 shows the totals for the top preference of subjects. Fig.9 shows the helpfulness of the different feedbacks to the task performance from the results of the 7-point questionnaires.

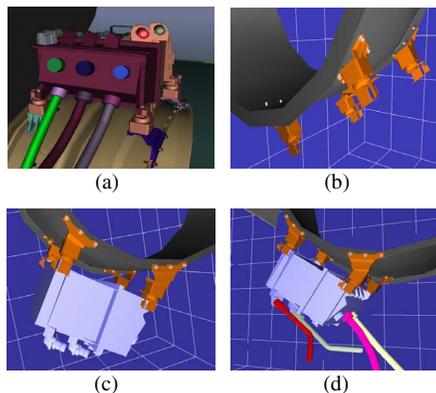


Fig. 7 Virtual Assembly Scenario of Sener Electronic Box Task

Conditions	First choice	Second choice	Third choice	Fourth choice
No feedback (neutral)	1	0	4	11
Visual	2	5	7	2
Auditory	3	9	2	2
Visual and Auditory	10	2	3	1

Table 4 Subjects Preference

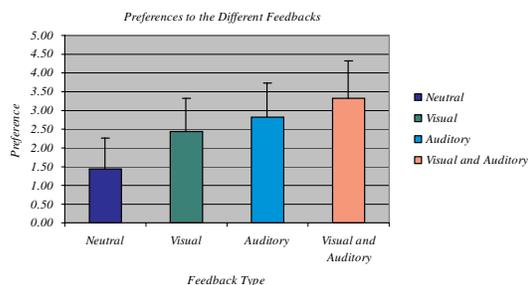


Fig. 8 Preferences to the Different Feedback

As is obvious from the data, the number of subjects preferring combined auditory plus visual feedback is statistically significantly larger than those preferring other feedback types. The number of subjects preferring the neutral feedback condition is obviously smaller than those preferring other feedback types.

From the subjects' general opinions and comments about their task completion experiences, more subjects prefer 3D localised to mono auditory feedback. In addition to these results, informal observation of frustration is much more frequent when subjects carry out the tasks under neutral feedback condition.

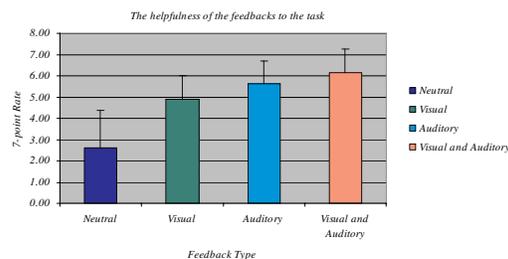


Fig. 9 Helpfulness of the Different Feedback to the Task Performance

4. Conclusion

A VAE system platform integrated with mono and 3D auditory feedback has been developed in order to explore and evaluate the effect of neutral, visual, auditory and integrated feedback mechanisms on the sense of presence and task performance in the context of assembly simulation. The task performance is measured on the basis of the time taken to complete the assembly task. The questionnaire is utilised to attain the subjective ranking among the four feedback conditions in order of preference when all trials are completed. The results of this research verified the original hypothesis that the performance is different between the four feedback conditions for the peg-in-a-hole task. Under the condition of the combined auditory and visual feedback, the assembly task performance is the best among the four feedback conditions. Under the neutral condition, the task completion time is the longest and the assembly task performance is the worst. For the subjective preference of the four different feedback conditions, the number of subjects preferring the combined auditory and visual feedback is statistically significantly larger than those preferring other feedback types. The number of subjects preferring the neutral feedback condition is obviously smaller than those preferring other feedback types. More subjects prefer the 3D localised auditory feedback to mono auditory feedback, but there is no obvious difference in the task completion time between the conditions of 3D localised and mono auditory feedback. The limitation of this research is that the peg-in-a-hole task case is a relatively simple one, even it has a common occurrence in assembly operations and it

integrates most of the assembly scenarios. For the future research, it required to determine how auditory feedback affects performance in specific design and tasks, and determine the substitution of 3D auditory feedback for force feedback in the assembly and manipulation tasks in virtual environments and how the 3D auditory feedback should be presented to maximise its utility.

5. References

- [1] F. Dai (Ed.), *Virtual Reality for Industrial Application*, Springer Verlag, Berlin, Heidelberg, 1998.
- [2] S. Jayaram, U. Jayaram, Y. Wang, H. Tirumali, K. Lyons and P. Hart, VADE: A Virtual Assembly Design Environment, IEEE CG & A, November 1999.
- [3] R. Steffan and T. Kuhlen, MAESTRO – A Tool for Interactive Assembly Simulation in Virtual Environments, *Proceedings of the joint IAO and EG workshop*, 16-18 May, 2001, Stuttgart, Germany.
- [4] D.R. Begault, *3D Sound for Virtual Reality and Multimedia*, Academic Press, Cambridge, Massachusetts, USA, 1994.
- [5] E.M. Wenzel, Localisation in Virtual Acoustic Displays, *Presence*, Vol. 1, No. 1, 80-107, Winter 1992.
- [6] K. Doel, P.G. Kry and D.K. Pai, Physically-based Sound Effects for Interactive Simulation and Animation, *Proceedings of ACM SIGGRAPH'2001*, 12-17 August 2001, Los Angeles, CA, USA.
- [7] J.F.O'Brien, P.R.Cook and G.Essl, Synthesising Sounds from Physically Based Motion, *Proceedings of ACM SIGGRAPH'2001*, Los Angeles, CA, USA.
- [8] J.K. Hahn, H.Fouad, L.Gritz and J.W. Lee, Integration Sounds and Motions in Virtual Environments, *Presence*, Vol.7 No.1, 67-77, February 1998.
- [9] Y.Kitamura, A. Yee and F. Kishino, A Sophisticated Manipulation Aid in a Virtual Environment using Dynamic Constraints among Object Faces, *Presence*, Vol.7, No.5, 460-477, October, 1998.
- [10] Y. Zhang, N. Murray and T. Fernando, Integration of 3D Sound Feedback into a Virtual Assembly Environment, Vol.1 of the *Proceedings of the 10th International Conference on Human-Computer Interaction (HCI International 2003)*, Crete, Greece, July 2003.
- [11] Y. Zhang and T. Fernando, 3D Auditory Feedback Act as Task Aid in a Virtual Assembly Environment, *Proceedings of the 21st Eurographics UK Chapter Conference (EGUK 2003)*, Birmingham, England, IEEE Computer Society Press, June 2003.